

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/00, 9/00	A1	(11) International Publication Number: WO 95/15375 (43) International Publication Date: 8 June 1995 (08.06.95)
(21) International Application Number: PCT/US94/10140 (22) International Filing Date: 7 September 1994 (07.09.94) (30) Priority Data: 08/162,809 3 December 1993 (03.12.93) US (71) Applicant: LA JOLLA CANCER RESEARCH FOUNDATION [US/US]; 10901 North Torrey Pines Road, La Jolla, CA 92037 (US). (72) Inventors: PASQUALE, Elena, B.; 6105 Dirac Street, San Diego, CA 92122 (US). SAJJADI, Fereydoun, G.; 1548 Plumtree Drive, Encinitas, CA 92024 (US). (74) Agents: IMBRA, Richard, J. et al.; Campbell and Flores, Suite 700, 4370 La Jolla Village Drive, San Diego, CA 92122 (US).	(81) Designated States: CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published With international search report.	
(54) Title: NOVEL Eph-RELATED TYROSINE KINASES, NUCLEOTIDE SEQUENCES AND METHODS OF USE (57) Abstract The invention provides substantially purified Eph-related protein tyrosine kinases, or functional fragments thereof, having about 23 to 66 percent amino acid sequence identity in their carboxyl terminal variable regions compared to known members of the Eph subclass of tyrosine kinases. Nucleic acids encoding such Eph-related protein tyrosine kinases, vectors and host cells are also provided. The invention also provides a method of diagnosing cancer and determining cancer prognosis. For example, the method provides for removing a tissue or cell sample from a subject suspected of having cancer and determining the level of Eph-related protein tyrosine kinase in the sample, wherein a change in the level or activity of an Eph-related protein tyrosine kinase compared to a normal sample indicates the presence of a cancer or indicates the level of malignancy of a cancer.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LJ	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

NOVEL Eph-RELATED TYROSINE KINASES, NUCLEOTIDE
SEQUENCES AND METHODS OF USE

This invention was funded in part by NIH Grants
HD 26351 and CA 56721. Accordingly, the United States
5 government has certain rights in the invention.

BACKGROUND OF THE INVENTION

This invention relates generally to protein
tyrosine kinases and, more particularly, to Eph-related
receptor tyrosine kinases and their manipulation for the
10 control of cellular processes.

Receptor tyrosine kinases comprise a large family
of proteins that share a number of structural features such
as a glycosylated extracellular ligand-binding domain, a
hydrophobic transmembrane domain and a conserved
15 cytoplasmic catalytic domain. Integral membrane tyrosine
kinases have been shown to mediate cellular signals
important for growth and differentiation. The transduction
of many extracellular signals to the cytoplasm occurs as a
result of the binding of ligands such as growth factors,
20 for example, to receptor tyrosine kinases at the cell
surface. In most cases, ligand binding activates the
cytoplasmic tyrosine kinase catalytic domain and culminates
in tyrosine phosphorylation of multiple substrates in the
cytoplasm.

25 Increased expression of membrane-spanning
receptor tyrosine kinases frequently has been associated
with alterations in normal cellular processes. The
affected cellular processes include cell proliferation,
differentiation and cancer, including, for example, human
30 cancers. Specific examples of such cancers can include
glioblastomas, squamous carcinomas and mammary carcinomas,
which are associated with the amplification of the EGF
receptor gene. Adenocarcinomas, breast cancers and gastric

cancers similarly are associated with aberrant expression of the HER2/neu receptor and certain breast carcinomas overexpress the erbB-3 gene, for example.

The correlation between aberrant expression and transforming ability also extends to members of the Eph subclass of receptor tyrosine kinases. For example, carcinomas of the liver, lung, breast and colon show elevated expression of Eph. Unlike many other tyrosine kinases, this elevated expression can occur in the absence of gene amplification or rearrangement. Such involvement of Eph in carcinogenesis also has been shown by the formation of foci of NIH 3T3 cells in soft agar and of tumors in nude mice following overexpression of Eph. Moreover, an antigen present on the surface of a pre-B cell leukemia cell line also has been identified as a member of the Eph subclass. Wicks et al., Proc. Natl. Acad. Sci., USA 89:1611-1615 (1992). This leukemia-specific marker, termed Hek, appears to be similar to the chicken Cek4 and mouse Mek4 of the Eph subclass of receptor tyrosine kinases (see Sajjadi et al., The New Biologist 3:769-778 (1991), which is incorporated herein by reference). As with Eph, Hek also was overexpressed in the absence of gene amplification or rearrangements in, for example, hemopoietic tumors and lymphoid tumor cell lines.

In addition to their roles in carcinogenesis, a number of transmembrane tyrosine kinases have been reported to play key roles during development. Examples include the mouse c-kit proto-oncogene and the *Drosophila* genes "sevenless" and "torso," which are involved in pattern formation. Consistent with this developmental role, many receptor tyrosine kinases other than those described above also have been shown to be developmentally regulated and predominantly expressed in embryonic tissues. Examples of these other tyrosine kinases include Cek1, which belongs to the FGF subclass, and the Cek4 and Cek5 tyrosine kinases

(Pasquale et al., Proc. Natl. Acad. Sci., USA 86:5449-5453 (1989); Sajjadi et al., *supra*, (1991); and Pasquale, E.B., Cell Reg. 2:523-534 (1991), all of which are incorporated herein by reference).

5 Eph was the first member of the Eph subclass of tyrosine kinases to be identified and characterized by molecular cloning (Hirai et al., Science 238:1717-1720 (1987)). The name Eph is derived from the name of the cell line from which the Eph cDNA was first isolated, the
10 erythropoietin-producing human hepatocellular carcinoma cell line, ETL-1. The general structure of Eph is similar to that of other receptor tyrosine kinases and consists of an extracellular domain, a single membrane spanning region and a conserved tyrosine kinase catalytic domain. However,
15 the structure of the extracellular domain of Eph, which comprises an immunoglobulin (Ig) domain at the amino terminus, followed by a cysteine-rich region and two fibronectin type III repeats in close proximity to the transmembrane domain, is completely distinct from that of
20 previously described receptor tyrosine kinases. The juxtamembrane domain and carboxy-terminus regions of Eph also are unrelated to the corresponding regions of other tyrosine kinase receptors. Thus, the discovery of Eph defined a new subclass of receptor-type tyrosine kinases.

25 In addition to the isolation and characterization of Eph, other related tyrosine kinases now have been identified. Cek4 and Cek5 were identified by screening a chicken embryo cDNA expression library with anti-phosphotyrosine antibodies (Sajjadi et al., *supra*, (1991)
30 and Pasquale, *supra*, (1991)). This method of identification was successful because Cek4 and Cek5 are expressed in embryonic tissues and have tyrosine kinase activity even when expressed as partial fragments in bacteria. Other Eph-related kinases that have been
35 identified include Hek (Wicks et al., *supra*, (1992)), Sek

(Gilardi-Hebenstreit et al, Oncogene 7:2499-2506 (1992)),
Eck (Lindberg and Hunter, Mol. Cell. Biol. 10:6316-6324
(1990)), Elk (Lhotak et al., Mol. Cell. Biol. 11:2496-2502
(1991)) and Eek (Chan and Watt, Oncogene 6:1057-1061
5 (1991)). These tyrosine kinases were cloned using a
variety of methods.

The number of existing Eph-related kinases is not
known and cannot be predicted. However, the Eph subclass
already represents the largest known subclass of receptor
10 tyrosine kinases, comprising at least 10 distinct members.
The kinases belonging to the Eph subclass are so classified
because each includes features such as the amino terminal
Ig domain, the cysteine-rich stretch and two fibronectin
type III repeats in the extracellular domain, which are
15 conserved within the Eph subclass. However, despite these
common structural features, the overall amino acid
sequences outside the catalytic domain are quite different,
indicating that different members of the Eph subclass
interact with distinct ligands and substrates and, thus,
20 exert distinct functions. This notion is supported by the
differential distribution of different Eph-related kinases
in adult tissues.

There is no indication whether other Eph-related
kinases exist and, if so, what their relationship is to the
25 known Eph-related kinases. Nevertheless, despite
similarities among the Eph-related receptor tyrosine
kinases, each is different and, as such, functions in
related but distinct cellular processes. For example,
many members of the Eph subclass are expressed in the
30 nervous system during development and thus are likely to be
involved in nerve regeneration processes. The aberrant
expression or uncontrolled regulation of any one of these
receptor tyrosine kinases can result in different
malignancies and pathological disorders. Therefore, the
35 identification and characterization of novel transmembrane

tyrosine kinases should provide important insights into the mechanisms underlying oncogenesis and cellular growth control pathways.

There thus exists a need to identify additional
5 receptor tyrosine kinases and to manipulate them in order to diagnose pathological conditions and control cellular processes. The present invention satisfies this need and provides related advantages as well.

SUMMARY

10 The invention is directed to substantially purified Eph-related protein tyrosine kinases, or functional fragments thereof, having about 23 to 66 percent amino acid sequence identity in their carboxyl terminal variable region compared to the other known members of the
15 Eph subclass of tyrosine kinases. Nucleic acids encoding such Eph-related protein tyrosine kinases, vectors and host cells also are provided. The invention also is directed to a method of diagnosing cancer. The method includes removing a tissue or cell sample from a subject suspected
20 of having cancer and determining the level of Eph-related protein tyrosine kinase in the sample, wherein a change in the level or activity of a Eph-related protein tyrosine kinase compared to a normal sample indicates the presence of a cancer or correlates with a specific prognosis.

25 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a comparison of the amino acid sequences from members of the Eph family. Dots replace residues in Cek4 (SEQ ID NO: 16), Cek6 (SEQ ID NO: 2), Cek7 (SEQ ID NO: 4), Cek8 (SEQ ID NO: 6), Cek9 (SEQ ID NO: 8),
30 Cek10 (SEQ ID NO: 10), Eck and Eph that are identical to the corresponding residue in Cek5 (SEQ ID NO: 18). Dashes represent gaps introduced in the sequences to aid in the

alignment. The insertion sequence of Cek5 also is presented (Cek5'; SEQ ID NO: 12) and the insertion sequences of Cek7' (SEQ ID NO: 20) and Cek10' (SEQ ID NO: 14) are in parentheses. The conserved cysteines are indicated by the symbol " and the kinase domain is delimited by arrows. Open circles indicate the hydrophobic and aromatic residues that are conserved in the first fibronectin type III repeat and asterisks indicate the conserved residues of the second fibronectin type III repeat. The filled circle indicates the site of putative tyrosine autophosphorylation in the catalytic domain. The putative signal peptide sequences and transmembrane domains are underlined. Amino acids are numbered at the left of the sequences. The symbol + indicates the location of the extracellular domain amino acid insertion RICTPDVSGTVGSRPAADH (SEQ ID NO: 23), corresponding to Cek6 amino acids 426-444. Alignments were made by eye in the regions corresponding to Cek5 residues 1-615 and using the program DFALIGN (Feng and Doolittle, J. Mol. Evol., 25:351-360 (1987), which is incorporated herein by reference) in the regions corresponding to Cek5 residues 616-995.

Figure 2 shows a RNA blot analysis of Cek mRNAs. Polyadenylated chicken RNA from 10-day embryonic and adult tissues was hybridized with Cek-specific cDNA probes and with a chicken β -actin probe. Hybridization conditions were as described in Example I. The positions of RNA molecular weight standards (in kilobases, kb) are indicated on the right. β -actin transcripts are present in the ~2.0 kb size range.

Figure 3 shows a RNA blot analysis of Cek5 mRNAs. Polyadenylated RNA from body tissues (lanes 1 and 2) and brain (lanes 3 and 4) of 10-day chicken embryos was hybridized with a Cek5-specific cDNA (lanes 1 and 3). The same blots were then stripped and rehybridized with a 48 bp oligonucleotide antisense probe corresponding to the

juxtamembrane insertion sequence of Cek5 (lanes 2 and 4). Hybridization conditions were as described in Example I. The positions of RNA molecular weight standards (in kb) are indicated on the right.

- 5 Figure 4 shows immunoblotting with antibodies to different Eph-related kinases. Fractions from 10-day embryonic brain containing either membrane-associated proteins (M) or soluble proteins (S) were probed with anti-Cek4 (4), anti-Cek8 (8,) or anti-Cek9 (9) antibodies.
- 10 Equal amounts of protein were loaded in all the lanes. IP, immunoprecipitates from 11-day embryonic retina with anti-Cek8 antibodies (8) or with normal rabbit IgGs (Ig). The immunoprecipitates were then probed with anti-Cek8 antibodies.
- 15 Figures 5.A. to 5.D. show the expression and tyrosine phosphorylation of Cek8 and Cek5 in transformed cell lines. Cell lysates were prepared from the rat central nervous system (CNS) tumor-derived cell lines B23, B28, B35, B49 and B50, the mouse embryonic carcinoma cell
- 20 line, P19, and the human keratinocyte cell line HaCaT (Ha). Panels A and B show immunoprecipitates with anti-Cek8 antibodies. Panels C and D show immunoprecipitates with anti-Cek5 antibodies. The immunoprecipitation was followed by *in vitro* kinase reaction in the samples shown in panel
- 25 C. The immunoblot in panel A was probed with anti-Cek8 antibodies. The immunoblots in panels B, C and D were probed with anti-phosphotyrosine antibodies.

- Figures 6.A. to 6.F. demonstrate that Cek8 phosphorylation on tyrosine is increased in transformed
- 30 cells and correlates with increased *in vitro* catalytic activity. Lysates from LMH cells and extracts of 10 day embryonic liver and adult liver were immunoprecipitated with anti-Cek8 antibodies, probed with anti-phosphotyrosine antibodies (panel A), then reprobed with anti-Cek8

antibodies (panel B). Lysates from normal chicken embryo fibroblasts and Rous sarcoma virus transformed chicken embryo fibroblasts were immunoprecipitated with anti-Cek8 antibodies, probed with anti-phosphotyrosine antibodies (panel C) and reprobed with anti-Cek8 antibodies (panel D).

Panels E and F show immunoblots of immunoprecipitated Cek8 (lane 1) or β -galactosidase-Cek4 fusion protein substrate (lanes 2-5). The fusion protein was phosphorylated for 1 min at 37 °C by Cek8 (lane 2), 1 min at 37 °C by tyrosine phosphorylated Cek8 (lane 3), 1 min at 0 °C by Cek8 (lane 4), 1 min at 0 °C by tyrosine phosphorylated Cek8 (lane 5). Immunoblots were probed with anti-phosphotyrosine antibodies (panel E) and reprobed with anti- β -galactosidase antibodies (panel F).

15

DETAILED DESCRIPTION OF THE INVENTION

The invention relates to the identification and characterization of seven novel members of the Eph subclass of membrane-spanning tyrosine kinases. The identification of these members doubles the number of kinases within this subclass, bringing the total to at least ten different Eph-related kinases. These Eph-related kinases therefore comprise the largest known subclass of integral membrane tyrosine kinases. The large number of different Eph-related kinases indicates that these receptors regulate a number of distinct cellular processes during development as well as in the adult organism. Therefore, identification of novel proteins within this subclass and isolation of their encoding nucleic acids allows the control of different cellular processes through the production of specific agonists and antagonists and through genetic therapy.

In one embodiment seven novel kinases of the Eph subclass of receptor protein tyrosine kinases have been

identified. The cDNAs encoding these Eph-related kinases were identified by hybridization at differential stringencies to identify distinct, but related receptor tyrosine kinases. All of the kinases exhibit gross structural features of known receptor tyrosine kinases in that they contain an extracellular ligand binding domain, a transmembrane domain and a cytoplasmic catalytic domain. These novel kinases are related to the Eph subclass of receptor tyrosine kinases and are designated Cek6 through Cek10* (SEQ ID NOS: 1 to 14, and 19 to 22.) The overall sequence identity between these Eph-related kinases varies significantly with each of the novel Eph-related receptors being identified by its carboxyl terminal variable region.

In another embodiment, the novel Eph-related kinases exhibit distinct tissue distribution patterns and developmental expression. Six of the kinases can be found to be expressed in both the embryonic brain and body tissues. The seventh Eph-related kinase, Cek5*, is expressed only in the embryonic brain. Indicative of their roles in cellular processes, such as embryonic signal transduction pathways, these Eph-related kinases display distinct patterns of expression in adult tissues, including the neuronal specific expression of Cek5*. These distinct patterns can be used to diagnose aberrations in normal cellular processes, such as those leading to uncontrolled malignant cell growth. For example, as described below, Cek8 activity is increased in various tumor cells as compared to normal cells. In addition to diagnosing such aberrations, it is also possible to treat defects caused by the unregulated expression of Eph-related kinases through the use of gene therapy. Reagents affecting the expression or activity of Eph-related kinases can also be useful for inducing nerve regeneration following injury.

As used herein, the term "Eph-related protein tyrosine kinase" or "Eph-related kinase" refers to a

receptor tyrosine kinase having an extracellular ligand binding domain, a transmembrane domain and a cytoplasmic catalytic domain, and belonging to the Eph subclass of receptor tyrosine kinases. Eph-related kinases include, for example, the receptor tyrosine kinases Cek6 (SEQ ID NO: 2), Cek7 (SEQ ID NO: 4), Cek7* (SEQ ID NO: 20), Cek7' (SEQ ID NO: 22), Cek8 (SEQ ID NO: 6), Cek9 (SEQ ID NO: 8), Cek10 (SEQ ID NO: 10), Cek5* (SEQ ID NO: 12) and Cek10* (SEQ ID NO: 14). Such kinases exhibit an overall amino acid sequence identity to Eph of greater than about 40 percent. The extreme carboxyl terminal cytoplasmic regions of the kinases are not well conserved and can be used to differentiate among them. This extreme carboxyl terminal cytoplasmic region begins just after the catalytic domain at about residue number 900 and extends to the C-terminal most residue. Therefore, the term "carboxyl terminal variable region" as used herein, refers to this extreme C-terminal region of the sequence which is divergent between the different members of the Eph subclass of tyrosine kinases. The actual sequence identities between different kinases within the Eph subclass are as follows: Cek5-Cek10: 66%; Cek5-Cek6: 54%; Cek5-Cek9: 50%; Cek5-Cek8: 38%; Cek5-Cek7: 34%; Cek5-Cek4: 24%; Cek5-Eek: 39%; Cek5-Eck: 36%; Cek5-Eph: 33%; Cek10-Cek6: 64%; Cek10-Cek9: 56%; Cek10-Cek8: 47%; Cek10-Cek7: 45%; Cek10-Cek4: 32%; Cek10-Eek: 41%; Cek10-Eck: 39%; Cek10-Eph: 37%; Cek6-Cek9: 46%; Cek6-Cek8: 50%; Cek6-Cek7: 40%; Cek6-Cek4: 31%; Cek6-Eek: 39%; Cek6-Eck: 36%; Cek6-Eph: 32%; Cek9-Cek8: 46%; Cek9-Cek7: 47%; Cek9-Cek4: 29%; Cek9-Eek: 36%; Cek9-Eck: 33%; Cek9-Eph: 35%; Cek8-Cek7: 37%; Cek8-Cek4: 26%; Cek8-Eek: 39%; Cek8-Eck: 36%; Cek8-Eph: 30%; Cek7-Cek4: 36%; Cek7-Eek: 35%; Cek7-Eck: 43%; Cek7-Eph: 37%; Cek4-Eek: 29%; Cek4-Eck: 27%; Cek4-Eph: 23%; Eek-Eck: 26%; Eek-Eph: 32%; Eck-Eph: 52%. Therefore, the carboxyl terminal variable region exhibits an amino acid sequence identity of about 23 to 66 percent between the different Eph-related kinases. The novel Eph-related kinases described herein fall within

this level of sequence divergence and can therefore be distinguished by comparison to the known members of the Eph subclass. Known members of this subclass include, for example, Eph, Cek4, Cek5, Mek4, Hek, Sek (or mouse Cek8),
5 Eck, Elk (or rat Cek6) and Eek.

It is understood that limited modifications may be made without destroying biological functions of Eph-related kinases and that only a portion of the entire primary structure may be required in order to effect a
10 particular activity. Such biological functions and activities can include, for example, signal transduction, ligand binding and/or tyrosine kinase activity. For example, the Eph-related kinases of the invention have amino acid sequences substantially similar to those shown
15 for Cek7, Cek7', Cek7'', Cek9, Cek10, Cek5*, Cek10* and chicken Cek6 and Cek8 in Figure 1 (hereinafter referred to as Cek6 through Cek10*), but minor modifications of these sequences which do not destroy their activity also fall within the definition of Eph-related kinases and within the
20 definition of the protein claimed as such. Moreover, fragments of the sequences of Cek6 through Cek10* in Figure 1 (SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 20 and 22), which retain the function of the entire protein as well as functional domains that contain at least one function of
25 the intact protein are included within the definition. Functional domains can include, for example, active ligand binding and catalytic domains. The boundaries of such domains are not important so long as activity is maintained. It is also understood that minor modifications
30 of the primary amino acid sequence can result in proteins which have substantially equivalent or enhanced function as compared to the sequences set forth in Figure 1 (SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 20 and 22). These modifications may be deliberate, as through site-directed
35 mutagenesis, or may be accidental such as through mutation in hosts which produce Eph-related kinases. All of these

modifications are included as long as biological function is retained. Further, various molecules can be attached to Eph-related kinases, for example, other proteins, carbohydrates, or lipids. Cek8 (SEQ ID NO: 6), for example, can contain complex N-linked oligosaccharides (see below). Such modifications are included within the definition of Eph-related tyrosine kinase.

The term "substantially purified," when used to describe the state of Eph-related tyrosine kinases denotes the protein free of a portion of the other proteins and molecules normally associated with or occurring with Eph-related kinases in their native environment. Such substantially purified Eph-related kinases can be derived from natural sources, recombinantly expressed or synthesized by *in vitro* methods so long as some portion of normally associated molecules is absent.

"Isolated" when used to describe the state of the nucleic acids encoding Eph-related tyrosine kinases denotes the nucleic acids free of at least a portion of the molecules associated with or occurring with Eph-related nucleic acids in the native environment.

As used herein, the term "vector" includes nucleic acids that are capable of harboring a natural or recombinant DNA sequence of interest. Vectors are usually derived from, or contain some sequences from, a natural source. For example, bacteriophage vectors containing specially engineered features that are largely derived from the phage's genome and are capable of carrying out some part of its infectious cycle. On the other hand, the sequences contained within plasmids are usually derived from different sources and compiled into a single molecule to carry out specific tasks. Thus, there are many different types of vectors and each is used according to the need to perform a desired function. Functions can include, for

example, propagation in a desired host, cloning recombinant or natural fragments of DNA, mutagenesis, expression and the like. In sum, "vector" is given a operative definition, and any DNA sequence which is capable of effecting a function of a specified DNA sequence disposed therein is included in this term as it is applied to the specified sequence.

The invention provides a substantially purified Eph-related protein tyrosine kinase, or functional fragment thereof. Also provided is a substantially purified chicken Eph-related protein tyrosine kinase. The substantially purified Eph-related protein tyrosine kinase exhibits about 23 to 66 percent amino acid sequence identity in its carboxyl terminal variable region compared to known members of the Eph subclass of tyrosine kinases. The amino acid sequences are substantially the same as that shown for Cek6 through Cek10* in Figure 1 (SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 20 and 22.)

The invention also provides an isolated nucleic acid encoding a Eph-related protein tyrosine kinase, or functional fragment thereof. The isolated nucleic acid encoding a Eph-related protein tyrosine kinase exhibits about 23 to 66 percent amino acid sequence identity in its carboxyl terminal variable region compared to known members of the Eph subclass of tyrosine kinases. The encoding nucleotide sequences are substantially the same as that shown for Cek6 (SEQ ID NO: 1), Cek7 (SEQ ID NO: 3), Cek8 (SEQ ID NO: 5), Cek9 (SEQ ID NO: 7), Cek10 (SEQ ID NO: 9), Cek5* (SEQ ID NO: 11), Cek10* (SEQ ID NO: 13), Cek7* (SEQ ID NO: 19) and Cek7' (SEQ ID NO: 21) (hereinafter Cek6 through Cek10*).

The isolation of seven cDNAs that encode novel Eph-related receptor tyrosine kinases is disclosed herein. The predicted amino acid sequences of these Eph-related

kinases are shown in Figure 1 along with other known Cek kinase sequences and those of Eph and Eck. A number of conserved features serve to define the newly discovered kinases as members of the Eph subclass. These include an amino terminal immunoglobulin domain followed by a cysteine-rich stretch in the extracellular domain, with the position of most cysteines conserved, and sequences corresponding to two fibronectin type III repeats in close proximity to the transmembrane domain (O'Bryan et al., Mol. Cell. Biol. 11:5016-5031 (1991) and Pasquale, *supra*, (1991), the former of which is incorporated herein by reference). Potential sites of N-glycosylation are primarily localized in the C-terminal half of the extracellular regions. The homologies in the extracellular domains indicates that the different members of the Eph family can bind a similar class of ligands. Figure 1 also shows that the Eph family, with the inclusion of the new members that have been identified, can now be considered the largest known family of membrane-spanning tyrosine kinases. Such a large number of tyrosine kinases in this one class is surprising in view of the fact that the other families of receptor tyrosine kinases have fewer members.

The catalytic domains of the Eph-related kinases are highly conserved and exhibit amino acid identities ranging between 61% and 90%. The C-terminal tails are less conserved (Figure 1) and therefore constitute a variable region which can be used to specify the distinct Eph-related kinases. Only one of the tyrosines in the C-terminal variable region, corresponding to tyrosine 939 of Cek5, is conserved in all the members of the Eph family, with the exception of Cek4. This conserved tyrosine residue represents a likely site of autophosphorylation and regulation, Ullrich and Schlessinger, Cell 61:203-212 (1990). The large size of the Eph subclass of receptor tyrosine kinases, the variability within their sequences and their different tissue distributions indicate that each

receptor can, for example, serve distinct functions during cellular processes.

The variability in both the lengths and sequences of the juxtamembrane domains observed in the Eph-related
5 kinases is unusual among tyrosine kinases belonging to the same subclass, Ullrich et al., *supra*, 1990. Because clones encoding variants with amino acid insertions in the juxtamembrane domain were isolated for Cek5, Cek7 and Cek10, the variability in the lengths of the juxtamembrane
10 domains is likely to originate by alternative splicing (Figure 1). Juxtamembrane domains are important for the modulation of receptor functions by heterologous stimuli, for example, through phosphorylation by other kinases. The juxtamembrane domains of the members of the Eph family
15 contain numerous serines, threonines and tyrosines that can serve as sites of regulation by phosphorylation, Kemp et al., Trends Biol. Sci. 15:342-346 (1990), which is incorporated herein by reference. For example, Cek9 and Cek10, as well as Cek5, Cek6, and Eck contain the consensus
20 sequence (S/T)P, which is recognized by proline-dependent protein kinases such as cdc2, Kemp et al., *supra*, (1990). Juxtamembrane domains have also been indicated to be important in the regulation of the subcellular distribution of the kinase and in the binding of some substrates
25 (Ullrich et al., *supra*, 1990).

The mRNA corresponding to Cek5* (SEQ ID NO: 11), the variant form of Cek5, was shown to be specifically expressed in the CNS, indicating that Cek5* functions primarily in neuronal cellular functions. Indicative of
30 this is another tyrosine kinase, src, which has been shown to encode neuronal specific variants containing 6 to 17 amino acid insertions in the regulatory (non-catalytic) region (Brugge et al., Nature 316:554-557 (1985); Martinez et al., Science 237:411-415 (1987); Pyper et al., Mol.
35 Cell. Biol. 10:2035-2040 (1990), all of which are

incorporated herein by reference). These neuronal forms of c-src have higher specific catalytic activity than non-neuronal c-src.

Although the predicted molecular masses of the
5 different members of the Eph family are similar, the sizes of their transcripts appear quite varied (4 to 10 kb). In addition, several mRNA species for each of the Eph-related kinases, particularly in the CNS, were detected using a panel of probes. As described below, the patterns of
10 expression of these novel Eph-related kinases are also distinct.

DNA sequences encoding the polypeptides of Eph-related kinases can be obtained by methods known to one skilled in the art. The sequences described herein are
15 sufficient for one skilled in the art to practice the invention. Such methods include, for example, cDNA synthesis and polymerase chain reaction (PCR). The need will determine which method or combination of methods is to be used to obtain the desired sequence. Expression can be
20 performed in any compatible vector/host system. Such systems include, for example, plasmids or phagemids in procaryotes such as *E. coli*, yeast systems and other eucaryotic systems such as mammalian cells. Additionally, the Eph-related kinases can also be expressed in soluble or
25 secreted form depending on the need and the vector/host system employed.

Such vectors and vector/host systems are known, or can be constructed by those skilled in the art and should contain all expression elements necessary for the
30 transcription, translation, regulation, and sorting of the polypeptide which makes up the Eph-related kinase. Other beneficial characteristics may also be contained within the vectors such as mechanisms for recovery of the nucleic acids in a different form. Phagemids are a specific

example of this because they can be used either as plasmids or as bacteriophage vectors. The vectors can also be for use in either procaryotic or eucaryotic host systems so long as the expression elements are of a compatible origin.

- 5 One of ordinary skill in the art will know which host systems are compatible with a particular vector. Thus, the invention provides vectors, host cells transformed with the vectors and Eph-related kinases produced from the host cells containing a nucleic acid encoding a Eph-related
10 kinase.

- The invention also provides methods of diagnosing cancer and determining cancer prognosis. The method includes removing a tissue or cell sample from a subject suspected of having cancer and determining the level of
15 Eph-related protein tyrosine kinase in said sample, wherein a change in the level or activity of a Eph-related protein tyrosine kinase compared to a normal sample indicates the presence of a cancer or indicates the level of malignancy of a cancer and, therefore, the most appropriate course of
20 treatment.

- As stated previously, receptor tyrosine kinases are involved in many signal transduction events that regulate important cellular processes. Such processes include, for example, cellular differentiation and
25 proliferation. Abnormal regulation or expression of the signal transduction machinery can lead to aberrant and malignant growth of the abnormally regulated cells. Abnormal expression of Eph is known to be associated with carcinomas of the liver, lung, breast and colon, for
30 example. Likewise, since some Eph-related tyrosine kinases are, at least, found within the same tissues as Eph, their abnormal expression may also lead to the development of the carcinomas described above as well as other types of cancers. For example, increased Cck8 activity was found in
35 embryonal carcinoma cells and a keratinocyte tumor cell

line (see Example II). Additionally, cancers of the neuronal lineage are likely to be caused by the abnormal expression or regulation of an Eph-related kinase such as Cek8 (see Example II) or Cek5* since this Eph-related kinase is found exclusively in neuronal tissues. Cek5*, Cek5 and the other Eph-related kinases expressed in the nervous system also are likely to be involved in nerve regeneration.

The important role that these receptor tyrosine kinases play in cellular processes can be advantageously used to diagnose early stages of cancer within a cell sample or tissue. A change in the amount or activity of an Eph-related kinase in a suspected sample, compared to a normal sample, will be indicative of cancerous stages and of their level of malignancy. Depending on whether the normal state is caused by the presence or absence of an Eph-related kinase, the change can involve either an increase or decrease in the amount or activity of the Eph-related kinase. For example, Cek8 activity is increased in various tumor cells (see Example II). Thus, increased activity of an Eph-related kinase of the invention such as Cek8 (SEQ ID NO: 6) can be useful for identifying the presence of transformed cells such as occur in a cancer.

One skilled in the art can measure the level or activity of an Eph-related kinase, for example, in a tissue sample obtained from a subject suspected of having a cancer or a developmental abnormality and the level or activity of the Eph-related kinase can be compared to the level or activity known to be present in a normal sample. Such a known level of activity can be determined by obtaining a significant number of tissue samples from subjects that do not have a cancer or a developmental abnormality and measuring the levels or activities of an Eph-related kinase in the population of samples. Methods for determining the level or activity of Eph-related kinases are known to the

skilled artisan and include, for example, RNA and protein blot analysis, ELISA using specific antibodies to each of the Eph-related kinases and direct measurement of catalytic activity such as tyrosine kinase activity. Such methods are described in detail in Example II or are otherwise known in the art (see, for example, Harlow et al., Antibodies: A Laboratory Manual Cold Spring Harbor Laboratory (1988), which is incorporated herein by reference).

The following examples are intended to illustrate, but not limit the invention.

EXAMPLE I

Isolation and Characterization of Eph-Related Tyrosine Kinases

This example shows the cloning and sequencing of the Eph-related kinases Cek6 through Cek10'. Structural characteristics and patterns of expression are also described.

To find novel members of the Eph family, various cDNA probes were used at different stringencies to screen a 10 day embryonic library as well as a 13 day embryonic brain cDNA library. The probes were derived from Cek4 (SEQ ID NO: 15) or Cek5 (SEQ ID NO: 17), which had been previously isolated based on phosphotyrosine content. Following subcloning and sequence analysis, it was found that the newly isolated cDNA clones encoded seven different Eph-related tyrosine kinases. Their isolation and structure are described below.

Briefly, a 10-day chicken embryo λ gt11 cDNA library (Clontech) and a 13-day embryonic brain λ gt11 cDNA library were used to isolate the cDNA clones. Screening was performed at different stringencies using the following procedure. Plaques were transferred to nylon membranes

(Micron Separations Inc.) on duplicate filters and hybridized to the appropriate probes at one of two stringencies (50% formamide, 42°C; or 50% formamide, 37°C). Conditions used were those recommended by the manufacturer and probes were detected using a nonradioactive DNA labeling and detection method (Boehringer Mannheim). Plaques identified as positive were subjected to three rounds of purification prior to DNA extraction using Lambda-TRAP (Clontech). Inserts from recombinant lambda DNA were subcloned in pBluescript vectors (Stratagene, San Diego, CA) using standard procedures and the sequences were analyzed on both strands, using the dideoxynucleotide chain-termination technique with Sequenase (United States Biochemical, Cleveland, OH).

Several clones distinguishable over known Eph tyrosine kinases were isolated using the Cek5 probe, which corresponded to nucleotides 495-3223 (Pasquale, *supra*, (1991)). The clones include: one Cek5' cDNA clone (from the chick embryo library); three Cek6 clones (two from the embryonic brain and one from the chick embryo library); one Cek7 clone (from the chick embryo library); one Cek7' clone (from the chick embryo library); one Cek7'' clone (from the embryonic brain library); one Cek9 clone (from the chick embryo library); one Cek10' clone (from the chick embryo library) and two Cek10 or Cek10' clones, which are indistinguishable because they do not encode the juxtamembrane domain, (one from the chick embryo and one from the embryonic brain library).

A Cek4 probe (corresponding to nucleotides 748-1756; see Sajjadi et al., *supra*, 1991), on the other hand, was used to isolate one Cek8 clone (from the chick embryo library). Also, following its initial isolation, a Cek10 probe, corresponding to residues 400-596 in Figure 2, was used to isolate clones extending further into the 5' end

from the chick embryo library. Of the two clones isolated, one represented Cek10 and one Cek10*.

The above-identified Eph-related kinases were characterized in terms of tissue distribution and expression by RNA blot analysis. Poly-A⁺ RNA was prepared from chicken tissues using the procedure of Badley et al., Biotechniques 6:114-116 (1988), which is incorporated herein by reference. Poly-A⁺ RNA (4-5 µg) was size-fractionated alongside RNA molecular weight markers on 0.9% agarose gels containing formaldehyde (Sambrook et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), which is incorporated herein by reference) and transferred to nitrocellulose filters (Schleicher & Schuell) according to methods known to one skilled in the art. The membranes were prehybridized for 2 hours and then hybridized under stringent conditions (50% formamide, 5x SSPE, 5x Denhardt's reagent, 0.5% SDS, 100 µg/ml salmon testes DNA, 42°C). Probes were labeled with ³²P dATP by the random-primed method of Feinberg and Vogelstein, Anal. Biochem. 132:6-13 (1983), which is incorporated herein by reference. T4 polynucleotide kinase was used to label the 5' end of the Cek5* specific oligonucleotide (Sambrook et al., *supra*, 1989). Filters were washed to a final stringency of 0.1x SSPE, 0.1% SDS at 58°C prior to exposure to Kodak XAR-5 X-ray film. For autoradiography of β-actin controls, intensifying screens were typically omitted and exposure time was reduced to 2 hours.

The following cDNA probes were used for RNA blot analysis: Cek4, 1.2 kb, same probe used for the library screening described previously, hybridizes to the region encoding amino acid residues 240-575; Cek5 probe, 1.2 kb, hybridizes to the 3' untranslated region; Cek6 5' probe, 1.3 kb, hybridizes to amino acid residues 1-438; Cek6 3' probe, 0.6 kb, hybridizes to the region following amino

acid 844; Cek7 5' probe, 0.4 kb, hybridizes to amino acid residues 1-136; Cek7 3' probe, 2.0 kb, hybridizes to the region following amino acid 137, including the 3' untranslated region; Cek8 probe, 1.2 kb, hybridizes to the region encoding amino acid residues 1-406; Cek9 probe, 0.6 kb, hybridizes to the region encoding amino acid residues 1-208; Cek10 probe, 0.6 kb, hybridizes to the region encoding the 10 C-terminal amino acids and to about 600 nucleotides of 3' untranslated region. For Cek6 and Cek7, the 3' Cek6 probe and the 5' Cek7 probe were used for the embryonic tissues mRNAs and a mixture of 5' and 3' probes for the adult tissues mRNAs.

Polyadenylated RNA was isolated from a number of adult chick tissues, as well as from brain and body tissues of 10-day embryos. These RNAs were then used for RNA blot analysis using the above specific probes. Probes were designed to minimize the possibility of cross-hybridization among the related kinases. Chicken β -actin DNA was used as a control probe (Cleveland et al., Cell 20:95-105 (1980), which is incorporated herein by reference).

The amino acid sequence of Cek4 (SEQ ID NO: 16) is 67% identical to that of Cek5 (SEQ ID NO: 18) in the catalytic and C-terminal regions and is most closely related to that of Cek7 (SEQ ID NO: 4) (75% amino acid identity in the same regions) (Figure 1). Preliminary data had indicated that Cek4 was highly expressed in the chicken developing brain and embryonic tissues, but no information was obtained on the adult pattern of expression in the chick. These data were therefore included in Figure 2. The 7.5 kb Cek4 transcript previously described was confirmed to be abundant in 10 day embryonic tissues. Expression was pronounced in the adult brain and retina, and lower but detectable in all other adult tissues examined, except the liver. In addition to the major 7.5 kb transcript, a smaller Cek4 transcript (of about 5 kb)

was found to be expressed at lower levels in the adult brain.

The Cek6 amino acid sequence (SEQ ID NO: 2) is most closely related to that of rat Elk (96% identity in the catalytic and C-terminal regions). Of the Cek members of the Eph subclass, Cek6 is most closely related to Cek5 (SEQ ID NO: 18) and Cek10 (SEQ ID NO: 10) (82% amino acid identity with both, in the catalytic and C-terminal regions) (Figure 1). The two Cek6 cDNAs that were isolated from a 13-day chick embryo brain library were identical and both encoded a protein with a deletion of 32 amino acids and an insertion of 19 amino acids in the extracellular region (Figure 1). However, these may be cloning artifacts, particularly the deletion, since it causes a shift in the reading frame and the premature termination of the encoded protein. A 4.4 kb Cek6 transcript was found to be expressed at high levels in the 10-days embryo and in adult brain, lung, heart and skeletal muscle (Figure 2). Low levels of Cek6 expression were detected in all other adult tissues tested. A second larger Cek6 transcript of about 6.5 kb was detected at low levels in the adult brain.

The amino acid sequence of Cek7 (SEQ ID NO: 4) is 71% identical to that of Cek5 (SEQ ID NO: 18) in the catalytic and C-terminal regions and is most closely related to those of Cek4 (SEQ ID NO: 16) and Cek9 (SEQ ID NO: 8) (75% amino acid identity with both, in the same regions) (Figure 1). A variant form of Cek7, containing a 22 amino acid insertion in the juxtamembrane domain (Figure 1) also was isolated and designated Cek7'. Cek7 (SEQ ID NO: 4) and Cek7' (SEQ ID NO: 20) may originate from the same gene by alternative splicing. A second variant form of Cek7, designated Cek7'' (SEQ ID NO: 22), which also presumably originates via alternative splicing, differs from Cek7 in the C-terminal 33 amino acids. Cek7 appears to have the lowest levels of expression among all the Eph

related kinases examined. Three different transcripts of about 4.4 kb, 7 kb and 8.5 kb were detected in the 10-day embryonic brain. Expression was weaker in the rest of the 10-day embryo, where only the 4.4 kb transcript could be
5 detected (Figure 2). Cek7 transcripts were not detected in the adult tissues, except for a barely detectable 8.5 kb transcript in the brain (Figure 2).

Cek8 (SEQ ID NO: 6) is equally related to Cek5 (SEQ ID NO: 18), Cek6 (SEQ ID NO: 2), Cek7 (SEQ ID NO: 4)
10 and Cek10 (SEQ ID NO: 10) (74% amino acid identity in the catalytic and C-terminal regions) (Figure 1). A single 6 kb Cek8 transcript was found to be present in both the 10-day embryonic brain and body tissues (Figure 2). Cek8 (SEQ ID NO: 6) expression appears to be the highest in adult
15 brain and retina and is also detectable in kidney, lung, skeletal muscle and thymus (Figure 2; see, also, Example II). Cek8 expression was not detected in heart and liver.

Cek9 (SEQ ID NO: 8) is most closely related to Cek5 (SEQ ID NO: 18) (77% identity at the amino acid level
20 in the catalytic and C-terminal regions (Figure 1). A 4.4 kb Cek9 transcript is present in embryonic brain and body tissues. Two additional and very minor transcripts of about 5.5 kb and 6.5 kb were detected exclusively in the 10-day embryonic brain (Figure 2). Among the adult tissues
25 examined, Cek9 expression is prominent in the thymus and detectable in brain, retina, kidney, lung and heart. None of the other kinases examined displays such an elevated level of expression in the thymus. Cek9 expression was not detected in skeletal muscle and liver.

30 Cek10 (SEQ ID NO: 10) is most closely related to Cek5 (SEQ ID NO: 18) and Cek6 (SEQ ID NO: 2) (84% amino acid identity with both in the catalytic and C-terminal regions) (Figure 1). A variant form of Cek10, containing a 15 amino acid insertion in the juxtamembrane domain

(Figure 1), was also isolated and designated Cek10* (SEQ ID NO: 14). Cek10 and Cek10* may originate from the same gene by alternative splicing. Northern blot analysis identified two Cek10 transcripts of about 4.4 kb and 6 kb, present at
5 different relative levels in 10-day embryonic brain and body tissues as well as in a number of adult tissues (Figure 2). Among the adult tissues examined, Cek10 expression was particularly prominent in the kidney. Lower
10 detectable transcripts were also present in brain, liver, heart, skeletal muscle and thymus.

A variant form of Cek5, containing a 16 amino acid insertion in the juxtamembrane domain, was also identified and termed Cek5* (SEQ ID NO: 12) (Figure 1).
15 This Cek5 variant may originate as a result of alternative splicing. With a Cek5 DNA probe recognizing both Cek5 and Cek5* (see Material and Methods), a 4.4 kb transcript was detected in both 10-day embryonic brain and body tissues (Figure 3, lanes 1 and 3). In addition, a much larger
20 transcript (of about 10 kb) was detected in the 10-day embryonic brain (Figure 3, lane 3). Consistently with the previously reported expression of the Cek5 protein, Cek5 transcripts are more abundant in the brain than in other 10-day embryonic tissues. Using a probe corresponding to
25 the 16 amino insertion in the juxtamembrane domain (Figure 3, lanes 2 and 4), Cek5* was found to be exclusively expressed in the CNS and only as the 4.4 kb transcript. Because Cek5 immunoreactivity in the CNS has been previously found to be confined to neurons, Cek5* appears to
30 be a neuronal specific variant of Cek5.

Polyclonal antibodies recognizing specifically Cek4, Cek8 and Cek9 have been obtained and will be used for the characterization of these kinases (see Example II). Peptides corresponding to the carboxy-terminal ends of
35 Cek4, Cek8 and Cek9 were coupled to bovine serum albumin

(BSA) with m-maleimido benzoyl-N-hydroxysuccinimide ester (Cek4) or with glutaraldehyde (Cek8 and Cek9) and used as immunogens. The peptides used were the following: Cek4, CLEHTTKNSPVPV (SEQ ID NO 24); Cek8, KMQQMHGRMVPV (SEQ ID NO 25) and Cek9, KVHLNQLEPVEV (SEQ ID NO 26). The carboxy-terminal regions were chosen because they are poorly conserved within the Eph subclass, increasing the likelihood of obtaining antibodies specific for each kinase.

10 The antibodies were purified from the antiserum by affinity-chromatography on the appropriate peptides coupled to N-hydroxy-succinimide-activated agarose (BioRad). As shown in Figure 4, after affinity purification the antibodies to Cek4, Cek8 and Cek9
15 recognize a single band of the expected apparent molecular mass (about 120 kiloDalton, kDa) in membranes-containing fractions isolated from 10-day embryonic brain, but not in fractions containing soluble proteins. These antibodies do not cross-react significantly with related members of the
20 Eph subclass (not shown) and can be used for different applications such as immunoblotting, immunofluorescence microscopy and immunoprecipitation (see Figure 4). All of the antibodies are capable of immunoprecipitating the kinases from tissue extracts and, as expected, the
25 immunoprecipitated kinases undergo *in vitro* autophosphorylation in the presence of ATP (see Example II).

 These techniques will allow the characterization of the kinases of the Eph subclass at the protein level.
30 Coupled to a solid support, the antibodies can also be used to purify the kinases from tissues and cell lines. In the cases tested, antibodies generated to the chicken Eph-related kinases recognize the corresponding mammalian homologues. Thus, these antibodies could be used, for

example, to screen tumor samples for the presence of the appropriate Eph-related kinases.

EXAMPLE II
CHARACTERIZATION OF CEK8

5 This example describes structural and functional characteristics of the Cek8 protein (SEQ ID NO: 6), including the expression and activity of Cek8 during development and in tumor cells.

A. Antibody preparation:

10 Cek8 expression and activity was examined using immunological and immunohistochemical methods. An antigen for raising anti-Cek8 antibodies was prepared by coupling the peptide KMQQMHGRMVPV (SEQ ID NO: 25), which consists of the eleven carboxy terminal amino acids of Cek8, including
15 an additional N-terminal lysine, to BSA using glutaraldehyde (Harlow and Lane, *supra*, 1988). An antigen for raising anti-Cek4 antibodies was prepared by coupling the peptide CLEHTKNSPVPV (SEQ ID NO: 24), which corresponds to the 12 carboxy terminal amino acids of Cek4,
20 including an additional cysteine at the N-terminus, to BSA using m-maleimidobenzoyl-N-hydroxysuccinimide ester (Harlow and Lane, *supra*, 1988). Anti-Cek5 antibodies and anti-phosphotyrosine antibodies were prepared as described by Pasquale, *supra*, (1991). Antisera were raised in
25 rabbits using standard methods (see, for example, Harlow and Lane, *supra*, 1988). The peptide antigen was coupled to N-hydroxy-succinimide-activated agarose and specific antisera were affinity purified.

B. Structural characterization of Cek8:

30 Cek8 was immunoprecipitated and examined by immunoblotting as described in Section C.1., below. The

affinity purified anti-Cek8 antibodies recognized a protein having an apparent molecular mass of about 120 kDa, which was the expected size for Cek8. The calculated molecular mass of Cek8, however, is less than the 120 kDa observed by SDS-PAGE. Since Cek8 contains three consensus sites of N-linked glycosylation, Cek8 was examined for such glycosylation. When chicken embryo fibroblasts were grown in the presence of 1.6 $\mu\text{g/ml}$ tunicamycin, which inhibits N-linked glycosylation, the apparent molecular mass of Cek8 decreased by about 10 kDa.

In order to characterize the carbohydrate moiety of Cek8, lectin affinity chromatography was performed. Ten day embryonic chicken brains were sonicated in 10 ml PBS containing protease inhibitors (protease inhibitors are 1 mM phenylmethylsulfonyl fluoride, 0.2 trypsin inhibitor units aprotinin/ml, 10 $\mu\text{g/ml}$ pepstatin and 10 $\mu\text{g/ml}$ leupeptin and 1 mM sodium orthovanadate, a phosphatase inhibitor. The sonicated material was centrifuged at 2000 x g for 5 min to remove insoluble material, then the supernatant was centrifuged at 200,000 x g for 40 min.

The pellet, which contained the membrane enriched fraction, was solubilized in PBS containing 0.1% Triton X-100. The solubilized sample was centrifuged 5 min in a microfuge and the supernatant was collected. The extract was dialyzed overnight at 4 °C against 10 mM Tris-HCl, pH 7.4, loaded onto various lectin columns, including concanavalin A, lentil lectin, wheat germ agglutinin, ricin I lectin, peanut lectin or *Ulex europaeus* I lectin (EY Laboratories, Inc.; San Mateo CA), and the columns were eluted with 0.1 M methyl α -D-mannopyranoside, 0.1 M D-mannose, 0.1 M N-acetyl-D-glucosamine, 0.1 M α -lactose, 0.1 M α -lactose or 0.05 M α -L-fucose, respectively. Fractions were collected and analyzed by immunoblotting for the presence of Cek8 as described below.

Cek8 bound to the concanavalin A, lentil lectin, ricin I and wheat germ agglutinin columns and was eluted with the appropriate buffers. These lectins preferentially recognize N-linked sugar chains. Thus, this result is in agreement with the observed inhibition of glycosylation by tunicamycin. In contrast, Cek8 does not bind to peanut lectin, which primarily recognizes O-linked chains.

Binding to concanavalin A and elution with the relatively low concentration of 0.1 M methyl α -D-mannopyranoside indicates that Cek8 contains biantennary complex type sugar chains (Osawa and Tsuji, Ann. Rev. Biochem. 56:21-42 (1987)). Binding to lentil lectin indicates that a fucose residue is present on the innermost N-acetylglucosamine residue in an oligosaccharide core. However, since Cek8 does not bind with *Ulex europaeus* I lectin, terminal fucose residues are not likely present (Sugii and Kabat, Carb. Res. 99:99-101 (1982)). Binding of Cek8 to wheat germ agglutinin indicates that sialic acid is present and binding to the ricin I column indicates that terminal β -galactosyl residues are present in complex sugar chains. These carbohydrate structures likely are located in the extracellular regions of Cek8 and can participate in interactions with extracellular molecules. In vivo phosphorylation on tyrosine of Cek8 can be achieved by exposing cells expressing Cek8 to wheat germ agglutinin.

C. Expression and Catalytic Activity of Cek8:

This section describes the methods for determining Cek8 expression and activity in various tissues during development and in tumor cells.

30 1. Methods

Cek8 expression and activity were determined by immunoprecipitation and immunoblot experiments. Cells from

90% confluent tissue culture plates were washed 3x with ice cold phosphate buffered saline (PBS), collected in cold RIPA buffer (150 mM sodium chloride, 10 mM sodium phosphate, pH 7.2, 1% deoxycholate, 1% Triton X-100, 0.1% SDS) containing protease inhibitors, and lysed by sonication. Phosphotyrosine was added to a final concentration of 8 mM when immunoblotting was performed using anti-phosphotyrosine antibodies.

Tissues were removed from adult chickens or chicken embryos and sonicated in PBS containing protease inhibitors. Whole embryos were collected and sonicated in PBS. Lysates were stored at -70 °C. Protein concentrations were determined using a Bio-Rad protein assay (Bio-Rad Laboratories; Richmond CA). For immunoprecipitations, tissue extracts were diluted in RIPA buffer. Cell lysates and tissue extracts in RIPA buffer were precleared using Staph A (Boehringer-Mannheim; Indianapolis IN) as described by Pasquale (*supra*. 1991). The samples then were incubated 40 min with 20 µg anti-Cek4, anti-Cek5 or anti-Cek8 antibodies or 20 µg control rabbit IgG preabsorbed to 20 µl Staph A. The amount of antibody was selected to ensure that all of the antigen in the extracts or lysates was precipitated.

Immunoprecipitated material was washed 3x with RIPA buffer and 1x with PBS. Sample buffer was added, the immunoprecipitates were boiled for 5 min, separated by SDS-PAGE on 7.5% gels and transferred to nitrocellulose as described by Towbin et al., Proc. Natl. Acad. Sci., USA 76:4350-4354 (1979), which is incorporated herein by reference. Following transfer, the filters were incubated overnight in Tris-hydroxyethylaminoethane-buffered saline (TBS) containing 3% BSA, then incubated 4 hr in 3% BSA containing 3 µg/ml anti-Cek4, anti-Cek5, anti-Cek8 or anti-phosphotyrosine antibody. The filters were rinsed with TBS, then incubated for 1 hr with 0.2 µg/ml protein A

peroxidase (Sigma; St. Louis MO) in TBS containing 3% BSA. The filters were rinsed several times with TBS and developed using enhanced chemiluminescence reagents (Amersham; Arlington Heights IL). In some experiments, after detection, the filters were dried for a few hours, then incubated in 3% BSA in TBS and probed with a different antibody.

In vitro phosphorylation was performed as described by Pasquale, *supra*, (1991). Briefly, Cek8 was immunoprecipitated from 10 day embryonic brain extracts or from cell lysates. In control experiments, Cek5 was immunoprecipitated. Immunoprecipitations were performed as described above. The immune complexes were incubated for 30 min at 37 °C in phosphorylation buffer (25 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid, pH 7.5, 10 mM MgCl₂, 10 mM MnCl₂, 1 mM sodium orthovanadate, 0.1% Triton X-100, 150 μM ATP). Sample buffer was added and electrophoresis and transfer to nitrocellulose were performed as described above. Following transfer, the filters were incubated overnight in 3% BSA in TBS, then incubated 4 hr in 3% BSA containing 3 μg/ml anti-phosphotyrosine antibodies.

2. Cek8 expression and activity during development

In whole embryo extracts, Cek8 expression was detectable at embryonic day 3, increased gradually between embryonic days 3 and 5, then remained relatively constant through embryonic day 10, which was the last timepoint examined. Cek8 was phosphorylated on tyrosine in vivo at a low level in 10 day embryonic brain. In addition, Cek8 underwent autophosphorylation on tyrosine in vitro in the presence of ATP and divalent metal ions.

Cek8 expression also was examined in various tissues of 10 day chicken embryos. Cek8 was most abundant

in the brain and retina, was expressed at substantial levels in thigh, gizzard and lungs and at lower levels in intestine, liver, lens and heart. Cek8 was not detectable in blood.

5 The developmental regulation of Cek8 expression was examined in greater detail in cerebrum, cerebellum, retina and thigh. In the cerebrum, Cek8 expression is low at embryonic day 6, then gradually increases to a maximal level at embryonic days 16 to 20. Cek8 expression is low,
10 but detectable, in adult cerebrum. In contrast, expression in the cerebellum is low at embryonic day 12 and barely detectable at later stages of development. In thigh muscle, Cek8 expression is highest at embryonic day 7, then decreases to barely detectable levels by day 13, before
15 terminal skeletal muscle differentiation occurs. In the retina, Cek8 expression remains relatively constant from embryonic day 8 until hatching.

Cek 8 expression also was examined by immunoperoxidase staining in chicken embryo frozen tissue
20 sections. Embryos were removed from eggs, fixed in 4% formaldehyde, 0.1 mM sodium orthovanadate in PBS for 16 to 24 hr, then cryoprotected in 20% sucrose in PBS, 0.1 mM sodium orthovanadate for 24 hr. Embryos were embedded in OCT compound (Miles Inc.; Tarrytown NY), then frozen in dry
25 ice/2-methylbutane. Ten μ m cryostat sections were collected on glass slides and stored at -70 °C.

The sections were treated with 0.3% hydrogen peroxide for 10 min, then blocked with 3% BSA or normal goat or horse serum in PBS for 30 min. Sections were
30 incubated with rabbit anti-Cek8 antibodies (10-20 μ g/ml) or mouse anti-200 kDa neurofilament protein antibodies (1 μ g/ml; Boehringer Mannheim; Indianapolis IN) in a 1:50 dilution of normal goat serum or horse serum for 30 min.

Controls were performed using anti-Cek8 antibodies that were preincubated with the antigen.

Following incubation with the primary antibody, the sections were rinsed with PBS and incubated with
5 biotinylated goat anti-rabbit or horse anti-mouse IgG (Vector Labs; Burlingame CA). After additional washes with PBS, the sections were incubated with an avidin-biotin-peroxidase complex or with an avidin-biotin-alkaline phosphatase complex (Vector Labs). Following several
10 washes in PBS, peroxidase or alkaline phosphatase were visualized using the appropriate substrate kit (Vector Labs). The sections then were rinsed in PBS, air dried, mounted in Permount and sealed with a #1 coverslip. Specimens were photographed with a Zeiss 405M inverted
15 microscope.

Cek8 immunoreactivity was intense in the spinal cord and the spinal nerves. Localization of Cek8 in the spinal nerves was similar to that of a 200 kDa neurofilament protein. At embryonic day 6, Cek8 expression
20 was restricted to the ventral portions of the spinal nerves, which contain axons of motor neurons.

The results of these experiments indicate that Cek8 is expressed early in development. In general, Cek8 expression is lower early in embryogenesis than at later
25 stages. Cek8 is differentially regulated in different tissues during development and expression is highest in the nervous system but also occurs in non-neuronal tissues. In view of these results, aberrant Cek8 expression or expression of an aberrant Cek8 protein can affect
30 development by causing defective signal transduction throughout an organism.

3. Cek8 expression and activity in tumor cells

Protein tyrosine kinase activity is tightly regulated in normal tissues and, in the tissues described above, Cek8 was phosphorylated on tyrosine at a low level.

5 It is well known that uncontrolled tyrosine kinase activity can lead to neoplastic transformation (Bishop, J.M., Cell 64:234-248 (1991)). Therefore, the expression and activation of Cek8 in a number of tumor cell lines was examined.

10 Because of the predominant expression of Cek8 in the brain and retina, Cek8 expression and activity was determined in a number of cell lines, B50, B49, B35, B28 and B23, which were derived from CNS system (CNS) tumors (Schubert et al., Nature 249:224-227 (1974), which is
15 incorporated herein by reference). B35 and B50 cells have neuronal properties and both expressed Cek8. However, Cek8 is substantially phosphorylated on tyrosine only in B50 cells (Figures 5.A. and 5.B.). B28 and B49 cells, which display glial characteristics, both expressed a moderate
20 level of Cek8 that is phosphorylated on tyrosine. B23 cells did not have detectable levels of Cek8.

The highest level of Cek8 expression was found in undifferentiated P19 embryonal carcinoma cells (McBurney and Rogers, J. Devel. Biol. 89:503-508 (1982), which is
25 incorporated herein by reference) and in HaCaT keratinocytes (Boukamp et al., J. Cell Biol. 106:761-770 (1988), which is incorporated herein by reference). In both of these cell lines, Cek8 was phosphorylated on tyrosine. Furthermore, comparable levels of Cek8
30 expression were observed in normal and Rous sarcoma virus-transformed chicken embryo fibroblasts. However, Cek8 was substantially phosphorylated on tyrosine only in the transformed cells (Figures 6.C. and 6.D.). In addition, in

LMH cells, which were derived from a hepatocellular carcinoma (Kawaguchi et al., Canc. Res. 47:4460-4464 (1987), which is incorporated herein by reference), Cek8 is highly phosphorylated on tyrosine as compared to adult or embryonic liver (Figures 6.A. and 6.B.).

For comparison, Cek5 expression and activation also was examined in the CNS tumor-derived cell lines. Cek5 was immunoprecipitated using anti-Cek5 antibodies followed by immunoblotting with anti-phosphotyrosine antibodies. Cek 5 was expressed in all of the cell lines derived from tumors of the CNS, with the highest expression in the B35 cells and the B49 cells. Tyrosine phosphorylation of Cek5 was observed in B28, B49 and B50 cells (Figures 5.C. and 5.D.). Cek5 also was highly expressed and phosphorylated in P19 cells and HaCaT cells. Thus, Cek5 expression and activation is similar, but not identical, to Cek8 expression and activation in tumor cells.

4. Effect of tyrosine phosphorylation on Cek8 kinase activity

The effect of tyrosine phosphorylation on the *in vitro* catalytic activity of Cek8 also was examined. *In vivo* substrates of Cek8 have not yet been identified. Therefore, a fusion protein consisting of the C-terminal 117 amino acids of Cek4 fused to β -galactosidase was used as an exogenous substrate. The fusion protein was purified from bacterial extracts by SDS-PAGE and eluted from the gel.

Assays were performed by incubating 1 μ g fusion protein substrate with Cek8 immunoprecipitate. Cek8 was immunoprecipitated from 10 day chick embryonic brain using 10 μ g anti-Cek8 antibodies and 5 μ l Staph A and was complexed to the antibodies and Staph A when the substrate was added. In some cases, Cek8 was phosphorylated for 1 hr

using the *in vitro* kinase reaction described above, in order to obtain Cek8 in a highly tyrosine-phosphorylated form. The fusion protein substrate then was added and phosphorylation of the substrate was allowed to proceed for 1 min at 0 °C or 37 °C in the phosphorylation buffer described above containing 200 μ M ATP. In parallel experiments, Cek8 that was not phosphorylated *in vitro* was used in the assay.

Following incubation, the samples were centrifuged briefly in a microfuge, 100 μ l 1% SDS in PBS was added to the pellets and the samples were heated at 95 °C for 5 min. The samples were centrifuged for 3 min and the supernatants were transferred to tubes containing anti- β -galactosidase antibodies bound to Staph A beads in 900 μ l RIPA buffer lacking SDS. Immunoprecipitation was performed as described above and the extent of tyrosine phosphorylation of the immunoprecipitated fusion protein substrate was analyzed by immunoblotting using anti-phosphotyrosine antibodies.

As shown in Figure 6.E., the phosphorylated form of Cek8 produced a greater amount of phosphorylation of the substrate on tyrosine at both 0 °C or 37 °C. These results indicate that activation of Cek8 by tyrosine phosphorylation increases the kinase activity of Cek8.

Although the invention has been described with reference to the disclosed embodiments, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: LA JOLLA CANCER RESEARCH FOUNDATION
- (ii) TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES, NUCLEOTIDE SEQUENCES, AND METHODS OF USE
- (iii) NUMBER OF SEQUENCES: 26
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: CAMPBELL AND FLORES
 - (B) STREET: 4370 La Jolla Village Drive, Suite 700
 - (C) CITY: San Diego
 - (D) STATE: California
 - (E) COUNTRY: United States of America
 - (F) ZIP: 92122
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE: 07-Sep-1994
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Imbra, Richard J.
 - (B) REGISTRATION NUMBER: 37,643
 - (C) REFERENCE/DOCKET NUMBER: FP-LJ 1114
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (619) 535-9001
 - (B) TELEFAX: (619) 535-8949

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3133 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: linear

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: join(3..419, 421..2858)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CA GAA ACC CTG ATG GAC ACA CGG ACA GCG ACG GCT GAG CTG GGC TGG	47
Glu Thr Leu Met Asp Thr Arg Thr Ala Thr Ala Glu Leu Gly Trp	
1 5 10 15	
ACT GCC AAC CCT CCG TCA GGG TGG GAA GAA GTG AGT GGC TAC GAC GAG	95
Thr Ala Asn Pro Pro Ser Gly Trp Glu Glu Val Ser Gly Tyr Asp Glu	
20 25 30	

AAC CTG AAC ACC ATC CGT ACC TAC CAG GTG TGC AAC GTC TTC GAG CCA Asn Leu Asn Thr 35 Ile Arg Thr Tyr Gln Val Cys Asn Val Phe Glu Pro 45	143
AAC CAG AAC AAC TGG CTC CTC ACC ACC TTC ATC AAC CGG CGC GGA GCC Asn Gln Asn Asn Trp Leu Leu Thr Thr Phe Ile Asn Arg Arg Gly Ala 60	191
CAC CGC ATC TAC ACT GAG ATG CGC TTC ACT GTG CGG GAC TGC AGC AGC His Arg Ile Tyr Thr Glu Met Arg Phe Thr Val Arg Asp Cys Ser Ser 75	239
CTC CCC AAC GTC CCC GGC TCC TGC AAG GAG ACC TTC AAC CTC TAC TAC Leu Pro Asn Val Pro Gly Ser Cys Lys Glu Thr Phe Asn Leu Tyr Tyr 95	287
TAT GAG ACA GAC TCT GTC ATT GCC ACT AAG AAG TCG GCC TTC TGG ACG Tyr Glu Thr Asp Ser Val Ile Ala Thr Lys Lys Ser Ala Phe Trp Thr 110	335
GAG GCA CCC TAC CTC AAA GTG GAC ACC ATT GCT GCT GAC GAG AGC TTT Glu Ala Pro Tyr Leu Lys Val Asp Thr Ile Ala Ala Asp Glu Ser Phe 125	383
TCC CAG GTG GAC TTT GGT GGC AGG TTG ATG AAG GGT T TTC TTC AAG Ser Gln Val Asp Phe Gly Gly Arg Leu Met Lys Gly Phe Phe Lys 140	429
AAG TGC CCA AGC GTG GTG CAG AAC TTC GCT ATC TTC CCT GAG ACG ATG Lys Cys Pro Ser Val Val Gln Asn Phe Ala Ile Phe Pro Glu Thr Met 155	477
ACG GGG GCA GAG AGC ACC TCT CTG GTG ACA GCA CGG GGC ACC TGC ATC Thr Gly Ala Glu Ser Thr Ser Leu Val Thr Ala Arg Gly Thr Cys Ile 170	525
CCC AAC GCT GAG GAG GTG GAC GTG CCC ATC AAG CTG TAC TGC AAC GGG Pro Asn Ala Glu Glu Val Asp Val Pro Ile Lys Leu Tyr Cys Asn Gly 190	573
GAT GGG GAG TGG ATG GTA CCC ATA GGT CGC TGC ACC TGC AAG GCT GGT Asp Gly Glu Trp Met Val Pro Ile Gly Arg Cys Thr Cys Lys Ala Gly 205	621
TAT GAG CCG GAA AAC AAC GTG GCT TGC AGA GCC TGC CCG GCT GGG ACA Tyr Glu Pro Glu Asn Asn Val Ala Cys Arg Ala Cys Pro Ala Gly Thr 220	669
TTC AAA GCC AGT CAG GGT GCG GGG CTG TGT GCC CGC TGT CCC CCC AAC Phe Lys Ala Ser Gln Gly Ala Gly Leu Cys Ala Arg Cys Pro Pro Asn 235	717
AGC CGC TCC AGC GCC GAG GCC TCA CCG CTC TGC GCC TGC CGC AAC GGC Ser Arg Ser Ser Ala Glu Ala Ser Pro Leu Cys Ala Cys Arg Asn Gly 250	765
TAC TTT CGG GCT GAC CTG GAC CCA CCG ACA GCT GCC TGC ACC AGC GTC Tyr Phe Arg Ala Asp Leu Asp Pro Pro Thr Ala Ala Cys Thr Ser Val 270	813
CCC TCT GGT CCA CGC AAC GTC ATC TCC ATT GTC AAT GAG ACC TCC ATC Pro Ser Gly Pro Arg Asn Val Ile Ser Ile Val Asn Glu Thr Ser Ile 285	861
ATC CTG GAG TGG AAC CCG CCA CGG GAG ACA GGA GGC CGG GAT GAT GTC Ile Leu Glu Trp Asn Pro Pro Arg Glu Thr Gly Gly Arg Asp Asp Val 300	909

ACT	TAC	AAC	ATT	GTC	TGC	AAG	AAG	TGC	CGG	GCA	GAC	CGG	CGT	GCC	TGC	957
Thr	Tyr	Asn	Ile	Val	Cys	Lys	Lys	Cys	Arg	Ala	Asp	Arg	Arg	Ala	Cys	
		305					310					315				
TCC	CGC	TGC	GAC	GAC	AAC	GTG	GAG	TTT	GTG	CCC	CGA	CAG	CTG	GGG	CTG	1005
Ser	Arg	Cys	Asp	Asp	Asn	Val	Glu	Phe	Val	Pro	Arg	Gln	Leu	Gly	Leu	
	320					325				330						
ACA	GAG	ACC	CGC	GTC	TTC	ATC	AGC	AGC	CTC	TGG	GCA	CAC	ACA	CCC	TAC	1053
Thr	Glu	Thr	Arg	Val	Phe	Ile	Ser	Ser	Leu	Trp	Ala	His	Thr	Pro	Tyr	
	335				340					345					350	
ACC	TTT	GAG	ATC	CAG	GCG	GTC	AAC	GGG	GTT	TCC	AAC	AAG	AGC	CCC	TTC	1101
Thr	Phe	Glu	Ile	Gln	Ala	Val	Asn	Gly	Val	Ser	Asn	Lys	Ser	Pro	Phe	
				355					360					365		
CCA	CCC	CAG	CAC	GTC	TCC	GTG	AAC	ATC	ACC	ACC	AAC	CAA	GCT	GCA	CCC	1149
Pro	Pro	Gln	His	Val	Ser	Val	Asn	Ile	Thr	Thr	Asn	Gln	Ala	Ala	Pro	
			370					375					380			
TCC	ACT	GTC	CCC	ATC	ATG	CAC	CAG	GTG	AGT	GCC	ACC	ATG	AGG	AGC	ATC	1197
Ser	Thr	Val	Pro	Ile	Met	His	Gln	Val	Ser	Ala	Thr	Met	Arg	Ser	Ile	
		385					390					395				
ACG	CTA	TCC	TGG	CCG	CAG	CCG	GAG	CAG	CCC	AAC	GGC	ATC	ATC	CTG	GAC	1245
Thr	Leu	Ser	Trp	Pro	Gln		Glu	Gln	Pro	Asn	Gly	Ile	Ile	Leu	Asp	
	400					405					410					
TAC	GAG	CTG	CGC	TAC	TAC	GAG	AAG	CTG	AGC	CGC	ATC	TGC	ACG	CCC	GAT	1293
Tyr	Glu	Leu	Arg	Tyr	Tyr	Glu	Lys	Leu	Ser	Arg	Ile	Cys	Thr	Pro	Asp	
	415				420					425					430	
GTC	AGC	GGC	ACT	GTG	GGC	TCG	AGA	CCG	GCG	GCG	GAC	CAC	AAC	GAG	TAC	1341
Val	Ser	Gly	Thr	Val	Gly	Ser	Arg	Pro	Ala	Ala	Asp	His	Asn	Glu	Tyr	
				435					440					445		
AAC	TCC	TCT	GTG	GCC	CGC	AGT	CAG	ACC	AAC	ACG	GCC	CGG	CTG	GAG	GGG	1389
Asn	Ser	Ser	Val	Ala	Arg	Ser	Gln	Thr	Asn	Thr	Ala	Arg	Leu	Glu	Gly	
			450					455					460			
CTG	CGC	CCT	GGC	ATG	GTG	TAC	GTG	GTG	CAG	GTG	CGA	GCA	AGG	ACG	GTG	1437
Leu	Arg	Pro	Gly	Met	Val	Tyr	Val	Val	Gln	Val	Arg	Ala	Arg	Thr	Val	
		465					470					475				
GCC	GGC	TAT	GGG	AAG	TAC	AGT	GGG	AAG	ATG	TGC	TTC	CAG	ACA	CTG	ACC	1485
Ala	Gly	Tyr	Gly	Lys	Tyr	Ser	Gly	Lys	Met	Cys	Phe	Gln	Thr	Leu	Thr	
	480					485					490					
GAT	GAT	GAC	TAC	AAG	TCT	GAG	CTG	AGG	GAG	CAG	CTG	CCA	TTG	ATT	GCG	1533
Asp	Asp	Asp	Tyr	Lys	Ser	Glu	Leu	Arg	Glu	Gln	Leu	Pro	Leu	Ile	Ala	
	495				500					505					510	
GGG	TCT	GCA	GCG	GCC	GGC	GTG	GTC	TTC	ATT	GTT	TCG	CTG	GTG	GCC	ATT	1581
Gly	Ser	Ala	Ala	Ala	Gly	Val	Val	Phe	Ile	Val	Ser	Leu	Val	Ala	Ile	
				515					520					525		
TCC	ATA	GTG	TGC	AGC	AGG	AAG	CGA	GCG	TAC	AGC	AAG	GAG	GTC	GTT	TAC	1629
Ser	Ile	Val	Cys	Ser	Arg	Lys	Arg	Ala	Tyr	Ser	Lys	Glu	Val	Val	Tyr	
		530						535					540			
AGC	GAT	AAG	CTG	CAG	CAC	TAC	AGC	ACC	GGG	AGA	GGG	TCT	CCG	GGA	ATG	1677
Ser	Asp	Lys	Leu	Gln	His	Tyr	Ser	Thr	Gly	Arg	Gly	Ser	Pro	Gly	Met	
		545					550					555				
AAG	ATT	TAC	ATC	GAC	CCC	TTC	ACT	TAT	GAG	GAC	CCC	AAC	GAG	GCA	GTG	1725
Lys	Ile	Tyr	Ile	Asp	Pro	Phe	Thr	Tyr	Glu	Asp	Pro	Asn	Glu	Ala	Val	
	560					565					570					

CGT GAG TTC GCC AAG GAG ATT GAC GTC TCC TTT GTG AAG ATT GAA GAG Arg Glu Phe Ala Lys Glu Ile Asp Val Ser Phe Val Lys Ile Glu Glu 575 580 585 590	1773
GTC ATT GGA GCA GGG GAG TTT GGA GAG GTG TAC AAA GGC CGC CTG AAG Val Ile Gly Ala Gly Glu Phe Gly Glu Val Tyr Lys Gly Arg Leu Lys 595 600 605	1821
TTG CCT GGC AAG CGG GAG ATC TAT GTG GCC ATC AAA ACA CTG AAG GCT Leu Pro Gly Lys Arg Glu Ile Tyr Val Ala Ile Lys Thr Leu Lys Ala 610 615 620	1869
GGC TAC TCA GAG AAG CAG CGC CGG GAT TTC CTG AGC GAA GCC AGC ATC Gly Tyr Ser Glu Lys Gln Arg Arg Asp Phe Leu Ser Glu Ala Ser Ile 625 630 635	1917
ATG GGG CAG TTT GAC CAC CCC AAC ATC ATC CGG CTG GAA GGG GTG GTG Met Gly Gln Phe Asp His Pro Asn Ile Ile Arg Leu Glu Gly Val Val 640 645 650	1965
ACC AAG AGC CGA CCA GTC ATG ATT ATC ACA GAG TTC ATG GAG AAT GGG Thr Lys Ser Arg Pro Val Met Ile Ile Thr Glu Phe Met Glu Asn Gly 655 660 665 670	2013
GCC CTG GAC TCG TTC CTG CGG CAA AAT GAT GGG CAG TTC ACA GTG ATC Ala Leu Asp Ser Phe Leu Arg Gln Asn Asp Gly Gln Phe Thr Val Ile 675 680 685	2061
CAG CTG GTG GGG ATG CTC AGA GGG ATT GCT GCT GGG ATG AAG TAC CTG Gln Leu Val Gly Met Leu Arg Gly Ile Ala Ala Gly Met Lys Tyr Leu 690 695 700	2109
GCA GAG ATG AAC TAT GTC CAC AGG GAT CTG GCG GCC AGG AAC ATT CTG Ala Glu Met Asn Tyr Val His Arg Asp Leu Ala Ala Arg Asn Ile Leu 705 710 715	2157
GTC AAC AGC AAC CTG GTG TGC AAA GTG TCA GAC TTT GGC CTC TCG CGC Val Asn Ser Asn Leu Val Cys Lys Val Ser Asp Phe Gly Leu Ser Arg 720 725 730	2205
TAC CTG CAG GAC GAC ACC TCT GAT CCC ACC TAC ACC AGC TCC TTG GGT Tyr Leu Gln Asp Asp Thr Ser Asp Pro Thr Tyr Thr Ser Ser Leu Gly 735 740 745 750	2253
GGG AAG ATC CCT GTG CGA TGG ACA GCA CCA GAG GCC ATT GCG TAC CGC Gly Lys Ile Pro Val Arg Trp Thr Ala Pro Glu Ala Ile Ala Tyr Arg 755 760 765	2301
AAG TTC ACG TCA GCC AGT GAC GTC TGG AGC TAT GGC ATC GTC ATG TGG Lys Phe Thr Ser Ala Ser Asp Val Trp Ser Tyr Gly Ile Val Met Trp 770 775 780	2349
GAG GTG ATG TCG TTC GGA GAG AGG CCC TAC TGG GAC ATG TCC AAC CAG Glu Val Met Ser Phe Gly Glu Arg Pro Tyr Trp Asp Met Ser Asn Gln 785 790 795	2397
GAC GTC ATC AAT GCC ATC GAG CAG GAC TAC CGG CTC CCG CCG CCC ATG Asp Val Ile Asn Ala Ile Glu Gln Asp Tyr Arg Leu Pro Pro Pro Met 800 805 810	2445
GAC TGC CCA GCT GCC CTG CAC CAA CTG ATG CTG GAC TGC TGG CAG AAG Asp Cys Pro Ala Ala Leu His Gln Leu Met Leu Asp Cys Trp Gln Lys 815 820 825 830	2493
GAC CGC AAC ACC CGG CCT CGC TTG GCC GAG ATT GTC AAC ACC CTG GAC Asp Arg Asn Thr Arg Pro Arg Leu Ala Glu Ile Val Asn Thr Leu Asp 835 840 845	2541

41

AAA ATG ATC CGC AAC CCG GCA AGC CTC AAA ACT GTG GCT ACC ATC ACC Lys Met Ile Arg Asn Pro Ala Ser Leu Lys Thr Val Ala Thr Ile Thr 850 855 860	2589
GCT GTG CCT TCT CAG CCC CTC CTC GAC CGC TCT ATC CCT GAT TTC ACT Ala Val Pro Ser Gln Pro Leu Leu Asp Arg Ser Ile Pro Asp Phe Thr 865 870 875	2637
GCC TTT ACC TCA GTA GAA GAC TGG CTG AGT GCC GTC AAG ATG AGC CAG Ala Phe Thr Ser Val Glu Asp Trp Leu Ser Ala Val Lys Met Ser Gln 880 885 890	2685
TAT AGA GAC AAC TTC CTG AGC GCT GGA TTC ACC TCC CTC CAG CTG GTC Tyr Arg Asp Asn Phe Leu Ser Ala Gly Phe Thr Ser Leu Gln Leu Val 895 900 905 910	2733
GCC CAG ATG ACA TCT GAA GAC CTC CTG AGA ATA GGA GTA ACG CTG GCT Ala Gln Met Thr Ser Glu Asp Leu Leu Arg Ile Gly Val Thr Leu Ala 915 920 925	2781
GGG CAC CAG AAG AAG ATC CTG AAC AGC ATC CAG TCC ATG CGC GTG CAG Gly His Gln Lys Lys Ile Leu Asn Ser Ile Gln Ser Met Arg Val Gln 930 935 940	2829
ATG AGT CAG TCT CCG ACC TCG ATG GCGTGACGTC CCTCGCTCGA CGAGGAGGGG Met Ser Gln Ser Pro Thr Ser Met Ala 945 950	2883
GACGGGGAGG GCAGGTGGCA GAGGTGGGAG GGGAGGAACT GATCTGATGG GAGCCGTGGG	2943
GCCGCAGCTG GAGAGGGGCA GCCACGGCCG GGGCTGTGCC TGACCGCGGA GGACGTTTCT	3003
GGGACTCGCC TCGGCCTGGT GACTTCCATC CCTCACCAAC AGAAGCACAC TTACCGATGT	3063
CACGGGGGAC AGCGTATAAA TAAGTATAAA TATGTACAAA TCATATATTT AAAAAAAAAA AAAAAAAAAG	3123 3133

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 951 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Glu Thr Leu Met Asp Thr Arg Thr Ala Thr Ala Glu Leu Gly Trp Thr 1 5 10 15
Ala Asn Pro Pro Ser Gly Trp Glu Glu Val Ser Gly Tyr Asp Glu Asn 20 25 30
Leu Asn Thr Ile Arg Thr Tyr Gln Val Cys Asn Val Phe Glu Pro Asn 35 40 45
Gln Asn Asn Trp Leu Leu Thr Thr Phe Ile Asn Arg Arg Gly Ala His 50 55 60
Arg Ile Tyr Thr Glu Met Arg Phe Thr Val Arg Asp Cys Ser Ser Leu 65 70 75 80
Pro Asn Val Pro Gly Ser Cys Lys Glu Thr Phe Asn Leu Tyr Tyr Tyr 85 90 95

Glu Thr Asp Ser Val Ile Ala Thr Lys Lys Ser Ala Phe Trp Thr Glu
 100 105 110
 Ala Pro Tyr Leu Lys Val Asp Thr Ile Ala Ala Asp Glu Ser Phe Ser
 115 120 125
 Gln Val Asp Phe Gly Gly Arg Leu Met Lys Gly Phe Phe Lys Lys Cys
 130 135 140
 Pro Ser Val Val Gln Asn Phe Ala Ile Phe Pro Glu Thr Met Thr Gly
 145 150 155 160
 Ala Glu Ser Thr Ser Leu Val Thr Ala Arg Gly Thr Cys Ile Pro Asn
 165 170 175
 Ala Glu Glu Val Asp Val Pro Ile Lys Leu Tyr Cys Asn Gly Asp Gly
 180 185 190
 Glu Trp Met Val Pro Ile Gly Arg Cys Thr Cys Lys Ala Gly Tyr Glu
 195 200 205
 Pro Glu Asn Asn Val Ala Cys Arg Ala Cys Pro Ala Gly Thr Phe Lys
 210 215 220
 Ala Ser Gln Gly Ala Gly Leu Cys Ala Arg Cys Pro Pro Asn Ser Arg
 225 230 235 240
 Ser Ser Ala Glu Ala Ser Pro Leu Cys Ala Cys Arg Asn Gly Tyr Phe
 245 250 255
 Arg Ala Asp Leu Asp Pro Pro Thr Ala Ala Cys Thr Ser Val Pro Ser
 260 265 270
 Gly Pro Arg Asn Val Ile Ser Ile Val Asn Glu Thr Ser Ile Ile Leu
 275 280 285
 Glu Trp Asn Pro Pro Arg Glu Thr Gly Gly Arg Asp Asp Val Thr Tyr
 290 295 300
 Asn Ile Val Cys Lys Lys Cys Arg Ala Asp Arg Arg Ala Cys Ser Arg
 305 310 315 320
 Cys Asp Asp Asn Val Glu Phe Val Pro Arg Gln Leu Gly Leu Thr Glu
 325 330 335
 Thr Arg Val Phe Ile Ser Ser Leu Trp Ala His Thr Pro Tyr Thr Phe
 340 345 350
 Glu Ile Gln Ala Val Asn Gly Val Ser Asn Lys Ser Pro Phe Pro Pro
 355 360 365
 Gln His Val Ser Val Asn Ile Thr Thr Asn Gln Ala Ala Pro Ser Thr
 370 375 380
 Val Pro Ile Met His Gln Val Ser Ala Thr Met Arg Ser Ile Thr Leu
 385 390 395 400
 Ser Trp Pro Gln Pro Glu Gln Pro Asn Gly Ile Ile Leu Asp Tyr Glu
 405 410 415
 Leu Arg Tyr Tyr Glu Lys Leu Ser Arg Ile Cys Thr Pro Asp Val Ser
 420 425 430
 Gly Thr Val Gly Ser Arg Pro Ala Ala Asp His Asn Glu Tyr Asn Ser
 435 440 445

Ser Val Ala Arg Ser Gln Thr Asn Thr Ala Arg Leu Glu Gly Leu Arg
 450 455 460
 Pro Gly Met Val Tyr Val Val Gln Val Arg Ala Arg Thr Val Ala Gly
 465 470 475 480
 Tyr Gly Lys Tyr Ser Gly Lys Met Cys Phe Gln Thr Leu Thr Asp Asp
 485 490 495
 Asp Tyr Lys Ser Glu Leu Arg Glu Gln Leu Pro Leu Ile Ala Gly Ser
 500 505 510
 Ala Ala Ala Gly Val Val Phe Ile Val Ser Leu Val Ala Ile Ser Ile
 515 520 525
 Val Cys Ser Arg Lys Arg Ala Tyr Ser Lys Glu Val Val Tyr Ser Asp
 530 535 540
 Lys Leu Gln His Tyr Ser Thr Gly Arg Gly Ser Pro Gly Met Lys Ile
 545 550 555 560
 Tyr Ile Asp Pro Phe Thr Tyr Glu Asp Pro Asn Glu Ala Val Arg Glu
 565 570 575
 Phe Ala Lys Glu Ile Asp Val Ser Phe Val Lys Ile Glu Glu Val Ile
 580 585 590
 Gly Ala Gly Glu Phe Gly Glu Val Tyr Lys Gly Arg Leu Lys Leu Pro
 595 600 605
 Gly Lys Arg Glu Ile Tyr Val Ala Ile Lys Thr Leu Lys Ala Gly Tyr
 610 615 620
 Ser Glu Lys Gln Arg Arg Asp Phe Leu Ser Glu Ala Ser Ile Met Gly
 625 630 635 640
 Gln Phe Asp His Pro Asn Ile Ile Arg Leu Glu Gly Val Val Thr Lys
 645 650 655
 Ser Arg Pro Val Met Ile Ile Thr Glu Phe Met Glu Asn Gly Ala Leu
 660 665 670
 Asp Ser Phe Leu Arg Gln Asn Asp Gly Gln Phe Thr Val Ile Gln Leu
 675 680 685
 Val Gly Met Leu Arg Gly Ile Ala Ala Gly Met Lys Tyr Leu Ala Glu
 690 695 700
 Met Asn Tyr Val His Arg Asp Leu Ala Ala Arg Asn Ile Leu Val Asn
 705 710 715 720
 Ser Asn Leu Val Cys Lys Val Ser Asp Phe Gly Leu Ser Arg Tyr Leu
 725 730 735
 Gln Asp Asp Thr Ser Asp Pro Thr Tyr Thr Ser Ser Leu Gly Gly Lys
 740 745 750
 Ile Pro Val Arg Trp Thr Ala Pro Glu Ala Ile Ala Tyr Arg Lys Phe
 755 760 765
 Thr Ser Ala Ser Asp Val Trp Ser Tyr Gly Ile Val Met Trp Glu Val
 770 775 780
 Met Ser Phe Gly Glu Arg Pro Tyr Trp Asp Met Ser Asn Gln Asp Val
 785 790 795 800

44

Ile Asn Ala Ile Glu Gln Asp Tyr Arg Leu Pro Pro Pro Met Asp Cys
805 810 815

Pro Ala Ala Leu His Gln Leu Met Leu Asp Cys Trp Gln Lys Asp Arg
820 825 830

Asn Thr Arg Pro Arg Leu Ala Glu Ile Val Asn Thr Leu Asp Lys Met
835 840 845

Ile Arg Asn Pro Ala Ser Leu Lys Thr Val Ala Thr Ile Thr Ala Val
850 855 860

Pro Ser Gln Pro Leu Leu Asp Arg Ser Ile Pro Asp Phe Thr Ala Phe
865 870 875 880

Thr Ser Val Glu Asp Trp Leu Ser Ala Val Lys Met Ser Gln Tyr Arg
885 890 895

Asp Asn Phe Leu Ser Ala Gly Phe Thr Ser Leu Gln Leu Val Ala Gln
900 905 910

Met Thr Ser Glu Asp Leu Leu Arg Ile Gly Val Thr Leu Ala Gly His
915 920 925

Gln Lys Lys Ile Leu Asn Ser Ile Gln Ser Met Arg Val Gln Met Ser
930 935 940

Gln Ser Pro Thr Ser Met Ala
945 950

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3059 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..2167

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

C CTC AAA TTC ACC CTG AGG GAC TGT AAC AGC CTT CCA GGA GGA CTT	46
Leu Lys Phe Thr Leu Arg Asp Cys Asn Ser Leu Pro Gly Gly Leu	
1 5 10 15	
GGG ACT TGC AAG GAG ACT TTT AAC ATG TAC TAC TTT GAG TCA GAT GAT	94
Gly Thr Cys Lys Glu Thr Phe Asn Met Tyr Tyr Phe Glu Ser Asp Asp	
20 25 30	
GAA GAT GGG AGG AAC ATC AGA GAG AAT CAG TAC ATC AAG ATA GAT ACC	142
Glu Asp Gly Arg Asn Ile Arg Glu Asn Gln Tyr Ile Lys Ile Asp Thr	
35 40 45	
ATT GCT GCT GAT GAG AGC TTC ACG GAG TTG GAC CTC GGC GAC AGA GTT	190
Ile Ala Ala Asp Glu Ser Phe Thr Glu Leu Asp Leu Gly Asp Arg Val	
50 55 60	
ATG AAG TTA AAC ACA GAA GTG AGA GAT GTT GGG CCT CTA ACA AAA AAA	238
Met Lys Leu Asn Thr Glu Val Arg Asp Val Gly Pro Leu Thr Lys Lys	
65 70 75	

45

GGA TTT TAC CTT GCT TTC CAG GAT GTG GGC GCC TGC ATT GCC CTG GTC	286
Gly Phe Tyr Leu Ala Phe Gln Asp Val Gly Ala Cys Ile Ala Leu Val	
80 85 90 95	
TCT GTG CGT GTG TAC TAC AAG AAA TGC CCA TCA GTG ATC CGC AAC CTG	334
Ser Val Arg Val Tyr Tyr Lys Lys Cys Pro Ser Val Ile Arg Asn Leu	
100 105 110	
GCA CGC TTT CCA GAT ACC ATC ACA GGA GCA GAT TCC TCG CAG CTG CTA	382
Ala Arg Phe Pro Asp Thr Ile Thr Gly Ala Asp Ser Ser Gln Leu Leu	
115 120 125	
GAA GTG TCA GGC GTC TGT GTC AAC CAC TCA GTG ACT GAT GAG GCA CCA	430
Glu Val Ser Gly Val Cys Val Asn His Ser Val Thr Asp Glu Ala Pro	
130 135 140	
AAG ATG CAC TGC AGT TCA GAG GGA GAA TGG CTG GTG CCC ATT GGG AAG	478
Lys Met His Cys Ser Ser Glu Gly Glu Trp Leu Val Pro Ile Gly Lys	
145 150 155	
TGT TTG TGC AAG GCA GGG TAC GAG GAG AAG AAC AAC ACC TGC CAA GCA	526
Cys Leu Cys Lys Ala Gly Tyr Glu Glu Lys Asn Asn Thr Cys Gln Ala	
160 165 170 175	
CCT TCT CCA GTC AGT AGT GTG AAA AAA GGG AAG ATA ACT AAA AAT AGC	574
Pro Ser Pro Val Ser Ser Val Lys Lys Gly Lys Ile Thr Lys Asn Ser	
180 185 190	
ATC TCC CTT TCC TGG CAG GAG CCA GAT CGA CCC AAC GGC ATC ATC CTG	622
Ile Ser Leu Ser Trp Gln Glu Pro Asp Arg Pro Asn Gly Ile Ile Leu	
195 200 205	
GAA TAC GAA ATC AAA TAT TTT GAA AAG GAC CAG GAG ACA AGC TAC ACC	670
Glu Tyr Glu Ile Lys Tyr Phe Glu Lys Asp Gln Glu Thr Ser Tyr Thr	
210 215 220	
ATC ATC AAA TCC AAA GAG ACC GCA ATT ACG GCA GAT GGC TTG AAA CCA	718
Ile Ile Lys Ser Lys Glu Thr Ala Ile Thr Ala Asp Gly Leu Lys Pro	
225 230 235	
GGC TCA GCG TAC GTC TTC CAG ATC CGA GCC CGG ACA GCT GCT GGC TAC	766
Gly Ser Ala Tyr Val Phe Gln Ile Arg Ala Arg Thr Ala Ala Gly Tyr	
240 245 250 255	
GGT GGC TTC AGT CGA AGA TTT GAG TTT GAA ACC AGC CCA GTG TTA GCT	814
Gly Gly Phe Ser Arg Arg Phe Glu Phe Glu Thr Ser Pro Val Leu Ala	
260 265 270	
GCA TCC AGT GAC CAG AGC CAG ATT CCT ATA ATT GTT GTG TCT GTA ACA	862
Ala Ser Ser Asp Gln Ser Gln Ile Pro Ile Ile Val Val Ser Val Thr	
275 280 285	
GTG GGA GTT ATT CTG CTG GCT GTT GTT ATC GGT TTC CTT CTC AGT GGA	910
Val Gly Val Ile Leu Leu Ala Val Val Ile Gly Phe Leu Leu Ser Gly	
290 295 300	
AGG CGC TGT GGC TAC AGC AAG GCT AAA CAA GAC CCA GAA GAA GAA AAG	958
Arg Arg Cys Gly Tyr Ser Lys Ala Lys Gln Asp Pro Glu Glu Glu Lys	
305 310 315	
ATG CAT TTT CAT AAT GGC CAC ATT AAA CTG CCT GGT GTA AGA ACC TAC	1006
Met His Phe His Asn Gly His Ile Lys Leu Pro Gly Val Arg Thr Tyr	
320 325 330 335	
ATT GAT CCC CAC ACC TAT GAG GAC CCT AAT CAA GCT GTC CAC GAG TTT	1054
Ile Asp Pro His Thr Tyr Glu Asp Pro Asn Gln Ala Val His Glu Phe	
340 345 350	

GCC AAG GAA ATA GAA GCT TCG TGC ATA ACC ATC GAG AGA GTT ATC GGA Ala Lys Glu Ile Glu Ala Ser Cys Ile Thr Ile Glu Arg Val Ile Gly 355 360 365	1102
GCT GGT GAA TTT GGA GAA GTC TGC AGT GGA CGG CTG AAA CTG CAG GGA Ala Gly Glu Phe Gly Glu Val Cys Ser Gly Arg Leu Lys Leu Gln Gly 370 375 380	1150
AAA CGC GAG TTT CCA GTG GCT ATC AAA ACC CTG AAG GTG GGC TAC ACA Lys Arg Glu Phe Pro Val Ala Ile Lys Thr Leu Lys Val Gly Tyr Thr 385 390 395	1198
GAG AAG CAA AGG CGA GAT TTC CTG GGA GAA GCG AGC ATC ATG GGG CAG Glu Lys Gln Arg Arg Asp Phe Leu Gly Glu Ala Ser Ile Met Gly Gln 400 405 410 415	1246
TTC GAC CAC CCC AAC ATC ATC CAC CTG GAA GGT GTC GTC ACA AAA AGC Phe Asp His Pro Asn Ile Ile His Leu Glu Gly Val Val Thr Lys Ser 420 425 430	1294
AAA CCT GTA ATG ATA GTA ACG GAA TAC ATG GAA AAT GGT TCT CTG GAT Lys Pro Val Met Ile Val Thr Glu Tyr Met Glu Asn Gly Ser Leu Asp 435 440 445	1342
ACA TTT TTA AAG AAG AAC GAT GGG CAG TTC ACG GTC ATT CAG CTG GTC Thr Phe Leu Lys Lys Asn Asp Gly Gln Phe Thr Val Ile Gln Leu Val 450 455 460	1390
GGG ATG CTG CGA GGC ATC GCA TCA GGG ATG AAG TAC CTG TCT GAC ATG Gly Met Leu Arg Gly Ile Ala Ser Gly Met Lys Tyr Leu Ser Asp Met 465 470 475	1438
GGT TAC GTA CAC AGA GAC CTC GCT GCC AGG AAT ATC CTC ATC AAC AGC Gly Tyr Val His Arg Asp Leu Ala Ala Arg Asn Ile Leu Ile Asn Ser 480 485 490 495	1486
AAC TTA GTC TGC AAG GTG TCT GAC TTT GGC CTC TCC AGA GTC CTA GAA Asn Leu Val Cys Lys Val Ser Asp Phe Gly Leu Ser Arg Val Leu Glu 500 505 510	1534
GAT GAT CCT GAA GCA GCG TAC ACA ACC AGG GGA GGG AAG ATC CCC ATC Asp Asp Pro Glu Ala Ala Tyr Thr Thr Arg Gly Gly Lys Ile Pro Ile 515 520 525	1582
CGA TGG ACG GCA CCT GAA GCA ATC GCC TTC CGC AAA TTC ACG TCG GCC Arg Trp Thr Ala Pro Glu Ala Ile Ala Phe Arg Lys Phe Thr Ser Ala 530 535 540	1630
AGC GAT GTG TGG AGC TAC GGC ATT GTG ATG TGG GAA GTG ATG TCC TAT Ser Asp Val Trp Ser Tyr Gly Ile Val Met Trp Glu Val Met Ser Tyr 545 550 555	1678
GGC GAG AGA CCT TAC TGG GAA ATG ACA AAC CAA GAT GTG ATT AAA GCC Gly Glu Arg Pro Tyr Trp Glu Met Thr Asn Gln Asp Val Ile Lys Ala 560 565 570 575	1726
GTG GAG GAA GGC TAT CGC CTG CCA AGT CCC ATG GAC TGC CCT GCT GCT Val Glu Glu Gly Tyr Arg Leu Pro Ser Pro Met Asp Cys Pro Ala Ala 580 585 590	1774
CTC TAC CAG TTG ATG CTT GAC TGC TGG CAG AAA GAC CGC AAC AGC AGG Leu Tyr Gln Leu Met Leu Asp Cys Gln Lys Asp Arg Asn Ser Arg 595 600 605	1822
CCC AAG TTT GAT GAA ATT GTC AGC ATG TTG GAC AAG CTC ATC CGT AAC Pro Lys Phe Asp Glu Ile Val Ser Met Leu Asp Lys Leu Ile Arg Asn 610 615 620	1870

CCA AGC AGC TTG AAG ACG TTG GTT AAT GCA TCG AGC AGA GTA TCA AAT Pro Ser Ser Leu Lys Thr Leu Val Asn Ala Ser Ser Arg Val Ser Asn 625 630 635	1918
TTG TTG GTA GAA CAC AGT CCA GTG GGG AGC GGT GCC TAC AGG TCA GTG Leu Leu Val Glu His Ser Pro Val Gly Ser Gly Ala Tyr Arg Ser Val 640 645 650 655	1966
GGT GAG TGG CTG GAA GCC ATC AAA ATG GGT CGA TAC ACC GAG ATT TTC Gly Glu Trp Leu Glu Ala Ile Lys Met Gly Arg Tyr Thr Glu Ile Phe 660 665 670	2014
ATG GAG AAT GGA TAC AGT TCG ATG GAT TCT GTG GCT CAG GTG ACC CTA Met Glu Asn Gly Tyr Ser Ser Met Asp Ser Val Ala Gln Val Thr Leu 675 680 685	2062
GAG GAT TTG AGG CGG CTG GGA GTG ACA CTT GTT GGT CAC CAG AAG AAG Glu Asp Leu Arg Arg Leu Gly Val Thr Leu Val Gly His Gln Lys Lys 690 695 700	2110
ATA ATG AAC AGC CTT CAA GAG ATG AAG GTC CAG TTG GTG AAT GGG ATG Ile Met Asn Ser Leu Gln Glu Met Lys Val Gln Leu Val Asn Gly Met 705 710 715	2158
GTG CCA TTG TAACTCGGTT TTTAAGTCAC TTCCTCGAGT GGTCGGTCCT Val Pro Leu 720	2207
GCACTTTGTA TACTAGCTCT GAGATTTATT TTGACTAAAG AAGAAAAAAG GGAAATTCAG	2267
TGGTTTCTGT AACTGAAGGA CGCTGGCTTC TGCCACAGCA TTTATAAAGC AGTGTTTGAC	2327
TGAAGTTTTC ATTTTCTTCC TATTTGTGTC CTCATTCTCA TGAAGTAAAT GTAACATGCA	2387
TGGAACATGG AAATGGATCT ACTGTACATG AGGTTACCCA ATTTCTTGCG CTTCAGCATG	2447
ACAACAGCAA GCCTTCCCAC CACATGTTGT CTATACATGG GAGATATATA TATATGCATA	2507
TATATATATA GCACCTTTAT ATACTGAATT ACAGCAGCAG CACATGTTAA TACTTCCAAG	2567
GACTTACTTG ACTAGAGAAG TTTTGCAGCC ATTGTGGGCT CACACAAGCT GCGGTTTACT	2627
GAAGTTTACT TCAAGTCTTA CTTGTCTACA GAAGTGTAAT GAAGAGCAAT ATGATTAGAT	2687
TATTTCTGGA TAGATATTTT GTTTTGTAAG TTTAAAAAAT CGTGTTACAC AGCGTTAAGT	2747
TATAGAGACT AGTGATATAA CATGTTGCTT GCTCAATGGC AAATACAATA CAGGGTGTAT	2807
ATTTTITTTCT CTCTGTGTTG CAAAGTTCTT TTAGTTTGCT CTTCTGTGAG GATAATACGT	2867
TATGATGTAT ATACTGTACA GTTTGCTACA CATCAGGTAC AAGATTGGGG CTTTCTCAAT	2927
GTTTTGTICT TTTTCCCTCT TTTGTTTCAT TTTGTCTTCC TTTTGTGTTA ACCACTATGC	2987
TTTGTATTTT TGCTGCTGTT TGGTTTGAGG CAACATATAA AGCTTTCAGG TGTTTTGATT	3047
ATAAAAAAAA AG	3059

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 722 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Leu Lys Phe Thr Leu Arg Asp Cys Asn Ser Leu Pro Gly Gly Leu Gly
 1           5           10           15

Thr Cys Lys Glu Thr Phe Asn Met Tyr Tyr Phe Glu Ser Asp Asp Glu
      20           25           30

Asp Gly Arg Asn Ile Arg Glu Asn Gln Tyr Ile Lys Ile Asp Thr Ile
      35           40           45

Ala Ala Asp Glu Ser Phe Thr Glu Leu Asp Leu Gly Asp Arg Val Met
      50           55           60

Lys Leu Asn Thr Glu Val Arg Asp Val Gly Pro Leu Thr Lys Lys Gly
      65           70           75           80

Phe Tyr Leu Ala Phe Gln Asp Val Gly Ala Cys Ile Ala Leu Val Ser
      85           90           95

Val Arg Val Tyr Tyr Lys Lys Cys Pro Ser Val Ile Arg Asn Leu Ala
      100          105          110

Arg Phe Pro Asp Thr Ile Thr Gly Ala Asp Ser Ser Gln Leu Leu Glu
      115          120          125

Val Ser Gly Val Cys Val Asn His Ser Val Thr Asp Glu Ala Pro Lys
      130          135          140

Met His Cys Ser Ser Glu Gly Glu Trp Leu Val Pro Ile Gly Lys Cys
      145          150          155          160

Leu Cys Lys Ala Gly Tyr Glu Glu Lys Asn Asn Thr Cys Gln Ala Pro
      165          170          175

Ser Pro Val Ser Ser Val Lys Lys Gly Lys Ile Thr Lys Asn Ser Ile
      180          185          190

Ser Leu Ser Trp Gln Glu Pro Asp Arg Pro Asn Gly Ile Ile Leu Glu
      195          200          205

Tyr Glu Ile Lys Tyr Phe Glu Lys Asp Gln Glu Thr Ser Tyr Thr Ile
      210          215          220

Ile Lys Ser Lys Glu Thr Ala Ile Thr Ala Asp Gly Leu Lys Pro Gly
      225          230          235          240

Ser Ala Tyr Val Phe Gln Ile Arg Ala Arg Thr Ala Ala Gly Tyr Gly
      245          250          255

Gly Phe Ser Arg Arg Phe Glu Phe Glu Thr Ser Pro Val Leu Ala Ala
      260          265          270

Ser Ser Asp Gln Ser Gln Ile Pro Ile Ile Val Val Ser Val Thr Val
      275          280          285

Gly Val Ile Leu Leu Ala Val Val Ile Gly Phe Leu Leu Ser Gly Arg
      290          295          300

```

Arg Cys Gly Tyr Ser Lys Ala Lys Gln Asp Pro Glu Glu Glu Lys Met
 305 310 315 320
 His Phe His Asn Gly His Ile Lys Leu Pro Gly Val Arg Thr Tyr Ile
 325 330 335
 Asp Pro His Thr Tyr Glu Asp Pro Asn Gln Ala Val His Glu Phe Ala
 340 345 350
 Lys Glu Ile Glu Ala Ser Cys Ile Thr Ile Glu Arg Val Ile Gly Ala
 355 360 365
 Gly Glu Phe Gly Glu Val Cys Ser Gly Arg Leu Lys Leu Gln Gly Lys
 370 375 380
 Arg Glu Phe Pro Val Ala Ile Lys Thr Leu Lys Val Gly Tyr Thr Glu
 385 390 395 400
 Lys Gln Arg Arg Asp Phe Leu Gly Glu Ala Ser Ile Met Gly Gln Phe
 405 410 415
 Asp His Pro Asn Ile Ile His Leu Glu Gly Val Val Thr Lys Ser Lys
 420 425 430
 Pro Val Met Ile Val Thr Glu Tyr Met Glu Asn Gly Ser Leu Asp Thr
 435 440 445
 Phe Leu Lys Lys Asn Asp Gly Gln Phe Thr Val Ile Gln Leu Val Gly
 450 455 460
 Met Leu Arg Gly Ile Ala Ser Gly Met Lys Tyr Leu Ser Asp Met Gly
 465 470 475 480
 Tyr Val His Arg Asp Leu Ala Ala Arg Asn Ile Leu Ile Asn Ser Asn
 485 490 495
 Leu Val Cys Lys Val Ser Asp Phe Gly Leu Ser Arg Val Leu Glu Asp
 500 505 510
 Asp Pro Glu Ala Ala Tyr Thr Thr Arg Gly Gly Lys Ile Pro Ile Arg
 515 520 525
 Trp Thr Ala Pro Glu Ala Ile Ala Phe Arg Lys Phe Thr Ser Ala Ser
 530 535 540
 Asp Val Trp Ser Tyr Gly Ile Val Met Trp Glu Val Met Ser Tyr Gly
 545 550 555 560
 Glu Arg Pro Tyr Trp Glu Met Thr Asn Gln Asp Val Ile Lys Ala Val
 565 570 575
 Glu Glu Gly Tyr Arg Leu Pro Ser Pro Met Asp Cys Pro Ala Ala Leu
 580 585 590
 Tyr Gln Leu Met Leu Asp Cys Trp Gln Lys Asp Arg Asn Ser Arg Pro
 595 600 605
 Lys Phe Asp Glu Ile Val Ser Met Leu Asp Lys Leu Ile Arg Asn Pro
 610 615 620
 Ser Ser Leu Lys Thr Leu Val Asn Ala Ser Ser Arg Val Ser Asn Leu
 625 630 635 640
 Leu Val Glu His Ser Pro Val Gly Ser Gly Ala Tyr Arg Ser Val Gly
 645 650 655

50

Glu Trp Leu Glu Ala Ile Lys Met Gly Arg Tyr Thr Glu Ile Phe Met
660 665 670

Glu Asn Gly Tyr Ser Ser Met Asp Ser Val Ala Gln Val Thr Leu Glu
675 680 685

Asp Leu Arg Arg Leu Gly Val Thr Leu Val Gly His Gln Lys Lys Ile
690 695 700

Met Asn Ser Leu Gln Glu Met Lys Val Gln Leu Val Asn Gly Met Val
705 710 715 720

Pro Leu

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2820 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 2..2548

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

C GGA GAG AGC CAG TTT GCC AAG ATT GAC ACC ATT GCT GCT GAT GAG Gly Glu Ser Gln Phe Ala Lys Ile Asp Thr Ile Ala Ala Asp Glu 1 5 10 15	46
AGC TTC ACC CAG GTG GAC ATT GGT GAC AGG ATC ATG AAG CTG AAT ACA Ser Phe Thr Gln Val Asp Ile Gly Asp Arg Ile Met Lys Leu Asn Thr 20 25 30	94
GAG GTG CCG GAC GTG GGG CCT CTC AGC AAG AAA GGG TTT TAC TTG GCT Glu Val Arg Asp Val Gly Pro Leu Ser Lys Lys Gly Phe Tyr Leu Ala 35 40 45	142
TTC CAG GAC GTC GGT GCC TGC ATT GCT TTG GTG TCT GTT CGT GTC TTC Phe Gln Asp Val Gly Ala Cys Ile Ala Leu Val Ser Val Arg Val Phe 50 55 60	190
TAT AAG AAG TGC CCA CTG ACA GTT CGA AAC CTG GCA CAG TTT CCA GAC Tyr Lys Lys Cys Pro Leu Thr Val Arg Asn Leu Ala Gln Phe Pro Asp 65 70 75	238
ACC ATT ACT GGG GCT GAT ACA TCC TCT CTG GTG GAG GTT CGT GGC TCC Thr Ile Thr Gly Ala Asp Thr Ser Ser Leu Val Glu Val Arg Gly Ser 80 85 90 95	286
TGT GTC AAC AAC TCG GAA GAG AAG GAC GTG CCA AAA ATG TAC TGC GGG Cys Val Asn Asn Ser Glu Glu Lys Asp Val Pro Lys Met Tyr Cys Gly 100 105 110	334
GCA GAT GGT GAA TGG CTG GTA CCC ATT GGC AAC TGT CTG TGC AAT GCT Ala Asp Gly Glu Trp Leu Val Pro Ile Gly Asn Cys Leu Cys Asn Ala 115 120 125	382
GGC TAT GAA GAA CGC AAT GGT GAA TGC CAA GCT TGC AAA ATC GGA TAC Gly Tyr Glu Glu Arg Asn Gly Glu Cys Gln Ala Cys Lys Ile Gly Tyr 130 135 140	430

TAC Tyr	AAG Lys	GCG Ala	CTC Leu	TCA Ser	ACA Thr	GAT Asp	GTT Val	GCA Ala	TGT Cys	GCC Ala	AAA Lys	TGC Cys	CCG Pro	CCT Pro	CAC His	478
145						150					155					
AGC Ser	TAC Tyr	TCC Ser	ATC Ile	TGG Trp	GAA Glu	GGC Gly	TCT Ser	ACC Thr	TCC Ser	TGC Thr	ACC Cys	TGT Cys	GAT Asp	CGG Arg	GGC Gly	526
160					165					170					175	
TTC Phe	TTC Phe	CGA Arg	GCA Ala	GAA Glu	AAT Asn	GAT Asp	GCT Ala	GCA Ala	TCC Ser	ATG Met	CCC Pro	TGC Cys	ACT Thr	CGC Arg	CCT Pro	574
				180					185					190		
CCA Pro	TCC Ser	GCA Ala	CCC Pro	CAG Gln	AAC Asn	CTG Leu	ATT Ile	TCC Ser	AAC Asn	GTC Val	AAC Asn	GAG Glu	ACG Thr	TCA Ser	GTG Val	622
			195					200					205			
AAC Asn	TTG Leu	GAG Glu	TGG Trp	AGC Ser	GCC Ala	CCA Pro	CAG Gln	AAC Asn	AAG Lys	GGA Gly	GGA Gly	CGG Arg	GAC Asp	GAC Asp	ATC Ile	670
	210						215					220				
TCC Ser	TAC Tyr	AAC Asn	GTG Val	GTG Val	TGC Cys	AAG Arg	CGC Cys	TGC Gly	GGG Ala	GCA Gly	GGG Glu	GAG Glu	CCC Pro	AGC Ser	CAC His	718
	225					230					235					
TGC Cys	CGG Arg	TCC Ser	TGT Cys	GGC Gly	AGT Ser	GGT Gly	GTA Val	CAT His	TTC Phe	AGC Ser	CCC Pro	CAG Gln	CAG Gln	AAC Asn	GGG Gly	766
240				245						250					255	
CTG Leu	AAA Lys	ACC Thr	ACG Thr	AAG Lys	GTT Val	TCC Ser	ATC Ile	ACT Thr	GAC Asp	CTC Leu	CTG Leu	GCA Ala	CAC His	ACC Thr	AAC Asn	814
				260					265					270		
TAC Tyr	ACC Thr	TTT Phe	GAG Glu	GTC Val	TGG Trp	GCA Ala	GTG Val	AAT Asn	GGA Gly	GTG Val	TCC Ser	AAG Lys	CAC His	AAC Asn	CCC Pro	862
			275					280					285			
AGC Ser	CAG Gln	GAC Asp	CAA Gln	GCT Ala	GTG Val	TCG Ser	GTC Val	ACT Thr	GTG Val	ACA Thr	ACT Thr	AAC Asn	CAA Gln	GCA Ala	GCT Ala	910
		290					295					300				
CCA Pro	TCC Ser	CCA Pro	ATT Ile	GCA Ala	TTG Leu	ATC Ile	CAG Gln	GCT Ala	AAA Lys	GAG Glu	ATA Ile	ACG Thr	AGG Arg	CAC His	AGC Ser	958
		305				310					315					
GTT Val	GCC Ala	TTG Leu	GCC Ala	TGG Trp	CTG Leu	GAA Glu	CCT Pro	GAC Asp	AGG Arg	CCC Pro	AAT Asn	GGA Gly	GTC Val	ATC Ile	CTG Leu	1006
320					325					330					335	
GAG Glu	TAC Tyr	GAA Glu	GTC Val	AAG Lys	TAC Tyr	TAC Tyr	GAA Glu	AAG Lys	GAC Asp	CAA Gln	AAC Asn	GAG Glu	CGC Arg	ACG Thr	TAT Tyr	1054
				340					345					350		
CGC Arg	ATT Ile	GTG Val	AAG Lys	ACA Thr	GCC Ala	TCC Ser	AGG Arg	AAT Asn	ACT Thr	GAC Asp	ATC Ile	AAA Lys	GGT Gly	TTG Leu	AAC Asn	1102
			355					360					365			
CCC Pro	CTG Leu	ACT Thr	TCA Ser	TAT Tyr	GTA Val	TTT Phe	CAT His	GTG Val	CGG Arg	GCC Ala	AGG Arg	ACA Thr	GCA Ala	GCA Ala	GGA Gly	1150
		370					375					380				
TAC Tyr	GGA Gly	GAC Asp	TTC Phe	AGT Ser	GGG Gly	CCG Pro	TTT Phe	GAG Glu	TTC Phe	ACA Thr	ACT Thr	AAC Asn	ACA Thr	GTT Val	CCT Pro	1198
	385					390					395					
TCC Ser	CCC Pro	ATC Ile	ATT Ile	GGC Gly	GAT Asp	GGT Gly	ACC Thr	AAT Asn	CCC Pro	ACA Thr	GTG Val	CTG Leu	CTT Leu	GTT Val	TCA Ser	1246
400					405					410					415	

GTG GCT GGC AGT GTT GTT CTT GTG GTC ATT CTC ATT GCA GCC TTT GTC Val Ala Gly Ser Val Val Leu Val Val Ile Leu Ile Ala Ala Phe Val 420 425 430	1294
ATC AGC AGG AGG CGC AGC AAA TAC AGT AAA GCT AAG CAA GAG GCA GAT Ile Ser Arg Arg Arg Ser Lys Tyr Ser Lys Ala Lys Gln Glu Ala Asp 435 440 445	1342
GAG GAG AAA CAT TTG AAC CAA GGT GTC AGA ACA TAT GTG GAT CCT TTT Glu Glu Lys His Leu Asn Gln Gly Val Arg Thr Tyr Val Asp Pro Phe 450 455 460	1390
ACA TAT GAG GAT CCA AAT CAA GCT GTG AGG GAA TTT GCC AAA GAA ATT Thr Tyr Glu Asp Pro Asn Gln Ala Val Arg Glu Phe Ala Lys Glu Ile 465 470 475	1438
GAT GCC TCC TGC ATA AAG ATT GAG AAA GTT ATT GGT GTG GGG GAA TTT Asp Ala Ser Cys Ile Lys Ile Glu Lys Val Ile Gly Val Gly Glu Phe 480 485 490 495	1486
GGT GAA GTA TGC AGT GGA CGT CTC AAA GTT CCA GGA AAA AGA GAA ATC Gly Glu Val Cys Ser Gly Arg Leu Lys Val Pro Gly Lys Arg Glu Ile 500 505 510	1534
TGT GTG GCT ATC AAG ACT CTG AAA GCT GGT TAC ACT GAC AAA CAA CGG Cys Val Ala Ile Lys Thr Leu Lys Ala Gly Tyr Thr Asp Lys Gln Arg 515 520 525	1582
AGA GAC TTC CTG AGT GAG GCC AGC ATC ATG GGA CAA TTT GAC CAC CCC Arg Asp Phe Leu Ser Glu Ala Ser Ile Met Gly Gln Phe Asp His Pro 530 535 540	1630
AAT ATC ATC CAC TTG GAA GGC GTT GTT ACT AAA TGT AAA CCA GTA ATG Asn Ile Ile His Leu Glu Gly Val Val Thr Lys Cys Lys Pro Val Met 545 550 555	1678
ATC ATA ACT GAG TAC ATG GAG AAT GGC TCC TTG GAT GCC TTC CTC CGG Ile Ile Thr Glu Tyr Met Glu Asn Gly Ser Leu Asp Ala Phe Leu Arg 560 565 570 575	1726
AAG AAT GAT GGC AGA TTT ACA GTA ATC CAG TTG GTG GGG ATG CTT CGT Lys Asn Asp Gly Arg Phe Thr Val Ile Gln Leu Val Gly Met Leu Arg 580 585 590	1774
GGC ATC GGC TCA GGA ATG AAG TAT CTG TCT GAC ATG AGC TAT GTG CAT Gly Ile Gly Ser Gly Met Lys Tyr Leu Ser Asp Met Ser Tyr Val His 595 600 605	1822
CGG GAT CTA GCT GCT CGA AAC ATA CTG GTC AAC AGC AAC TTG GTC TGC Arg Asp Leu Ala Ala Arg Asn Ile Leu Val Asn Ser Asn Leu Val Cys 610 615 620	1870
AAA GTG TCT GAC TTT GGC ATG TCC CGT GTC CTG GAA GAT GAC CCT GAG Lys Val Ser Asp Phe Gly Met Ser Arg Val Leu Glu Asp Asp Pro Glu 625 630 635	1918
GCA GCT TAT ACC ACA CGG GGT GGC AAG ATC CCT ATC CGA TGG ACT GCA Ala Ala Tyr Thr Thr Arg Gly Gly Lys Ile Pro Ile Arg Trp Thr Ala 640 645 650 655	1966
CCA GAG GCA ATT GCC TAC CGT AAA TTT ACA TCG GCT AGT GAC GTG TGG Pro Glu Ala Ile Ala Tyr Arg Lys Phe Thr Ser Ala Ser Asp Val Trp 660 665 670	2014
AGC TAT GGC ATC GTC ATG TGG GAA GTG ATG TCC TAT GGA GAG AGA CCT Ser Tyr Gly Ile Val Met Trp Glu Val Met Ser Tyr Gly Glu Arg Pro 675 680 685	2062

53

TAC TGG GAT ATG TCC AAT CAA GAC GTT ATT AAA GCC ATT GAG GAA GGG Tyr Trp Asp Met Ser Asn Gln Asp Val Ile Lys Ala Ile Glu Glu Gly 690 695 700	2110
TAT CGG TTG CCA CCC CCA ATG GAC TGC CCC ATT GCT CTC CAT CAG CTG Tyr Arg Leu Pro Pro Pro Met Asp Cys Pro Ile Ala Leu His Gln Leu 705 710 715	2158
ATG TTA GAC TGC TGG CAG AAG GAA CGC AGC GAC AGA CCT AAA TTT GGA Met Leu Asp Cys Trp Gln Lys Glu Arg Ser Asp Arg Pro Lys Phe Gly 720 725 730 735	2206
CAG ATT GTC AAC ATG CTG GAC AAA CTC ATC CGC AAC CCT AAC AGC CTG Gln Ile Val Asn Met Leu Asp Lys Leu Ile Arg Asn Pro Asn Ser Leu 740 745 750	2254
AAG AGG ACA GGC AGC GAG AGC TCC AGA CCC AGC ACA GCC CTG CTG GAT Lys Arg Thr Gly Ser Glu Ser Ser Arg Pro Ser Thr Ala Leu Leu Asp 755 760 765	2302
CCC AGC TCC CCG GAG TTC TCG GCG GTT GTT TCT GTC AGT GAC TGG CTC Pro Ser Ser Pro Glu Phe Ser Ala Val Val Ser Val Ser Asp Trp Leu 770 775 780	2350
CAA GCC ATT AAA ATG GAG CGA TAC AAG GAT AAC TTC ACA GCT GCT GGC Gln Ala Ile Lys Met Glu Arg Tyr Lys Asp Asn Phe Thr Ala Ala Gly 785 790 795	2398
TAT ACC ACC CTA GAG GCT GTG GTG CAT ATG AAC CAG GAC GAC CTG GCC Tyr Thr Thr Leu Glu Ala Val Val His Met Asn Gln Asp Asp Leu Ala 800 805 810 815	2446
AGG ATC GGG ATC ACT GCC ATC ACA CAC CAG AAC AAG ATC TTG AGC AGC Arg Ile Gly Ile Thr Ala Ile Thr His Gln Asn Lys Ile Leu Ser Ser 820 825 830	2494
GTT CAA GCC ATG CGC AGC CAA ATG CAA CAG ATG CAC GGC AGG ATG GTG Val Gln Ala Met Arg Ser Gln Met Gln Gln Met His Gly Arg Met Val 835 840 845	2542
CCC GTC TGAGCCAGTA CTGAATAAAC TCAAACTCT TGAAATTAGT TTACCTCATC Pro Val	2598
CATGCACTTT AATTGAAGAA CTGCACTTTT TTTACTTCGT CTCCTCGCCC GTTGAAATAA	2658
AGATCTGCAG CATTGCTTGA TGTACAGATT GTGGAAACCG AGCGTGTGTT GGGAGGGGGG	2718
CCTCCAGAAA TGACAAGCCG TCATTTTAAA CCAGACCTGG AACAAATTGT TTCTTGAAC	2778
ATACTTCTCT GTTGATCAAC GATATGTAAA ATACATGTAT CC	2820

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 849 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Gly	Glu	Ser	Gln	Phe	Ala	Lys	Ile	Asp	Thr	Ile	Ala	Ala	Asp	Glu	Ser
1					5				10					15	

54

Phe Thr Gln Val Asp Ile Gly Asp Arg Ile Met Lys Leu Asn Thr Glu
 20 25 30
 Val Arg Asp Val Gly Pro Leu Ser Lys Lys Gly Phe Tyr Leu Ala Phe
 35 40 45
 Gln Asp Val Gly Ala Cys Ile Ala Leu Val Ser Val Arg Val Phe Tyr
 50 55 60
 Lys Lys Cys Pro Leu Thr Val Arg Asn Leu Ala Gln Phe Pro Asp Thr
 65 70 75 80
 Ile Thr Gly Ala Asp Thr Ser Ser Leu Val Glu Val Arg Gly Ser Cys
 85 90 95
 Val Asn Asn Ser Glu Glu Lys Asp Val Pro Lys Met Tyr Cys Gly Ala
 100 105 110
 Asp Gly Glu Trp Leu Val Pro Ile Gly Asn Cys Leu Cys Asn Ala Gly
 115 120 125
 Tyr Glu Glu Arg Asn Gly Glu Cys Gln Ala Cys Lys Ile Gly Tyr Tyr
 130 135 140
 Lys Ala Leu Ser Thr Asp Val Ala Cys Ala Lys Cys Pro Pro His Ser
 145 150 155 160
 Tyr Ser Ile Trp Glu Gly Ser Thr Ser Cys Thr Cys Asp Arg Gly Phe
 165 170 175
 Phe Arg Ala Glu Asn Asp Ala Ala Ser Met Pro Cys Thr Arg Pro Pro
 180 185 190
 Ser Ala Pro Gln Asn Leu Ile Ser Asn Val Asn Glu Thr Ser Val Asn
 195 200 205
 Leu Glu Trp Ser Ala Pro Gln Asn Lys Gly Gly Arg Asp Asp Ile Ser
 210 215 220
 Tyr Asn Val Val Cys Lys Arg Cys Gly Ala Gly Glu Pro Ser His Cys
 225 230 235 240
 Arg Ser Cys Gly Ser Gly Val His Phe Ser Pro Gln Gln Asn Gly Leu
 245 250 255
 Lys Thr Thr Lys Val Ser Ile Thr Asp Leu Leu Ala His Thr Asn Tyr
 260 265 270
 Thr Phe Glu Val Trp Ala Val Asn Gly Val Ser Lys His Asn Pro Ser
 275 280 285
 Gln Asp Gln Ala Val Ser Val Thr Val Thr Thr Asn Gln Ala Ala Pro
 290 295 300
 Ser Pro Ile Ala Leu Ile Gln Ala Lys Glu Ile Thr Arg His Ser Val
 305 310 315 320
 Ala Leu Ala Trp Leu Glu Pro Asp Arg Pro Asn Gly Val Ile Leu Glu
 325 330 335
 Tyr Glu Val Lys Tyr Tyr Glu Lys Asp Gln Asn Glu Arg Thr Tyr Arg
 340 345 350
 Ile Val Lys Thr Ala Ser Arg Asn Thr Asp Ile Lys Gly Leu Asn Pro
 355 360 365

Leu Thr Ser Tyr Val Phe His Val Arg Ala Arg Thr Ala Ala Gly Tyr
 370 375 380
 Gly Asp Phe Ser Gly Pro Phe Glu Phe Thr Thr Asn Thr Val Pro Ser
 385 390 395 400
 Pro Ile Ile Gly Asp Gly Thr Asn Pro Thr Val Leu Leu Val Ser Val
 405 410 415
 Ala Gly Ser Val Val Leu Val Val Ile Leu Ile Ala Ala Phe Val Ile
 420 425 430
 Ser Arg Arg Arg Ser Lys Tyr Ser Lys Ala Lys Gln Glu Ala Asp Glu
 435 440 445
 Glu Lys His Leu Asn Gln Gly Val Arg Thr Tyr Val Asp Pro Phe Thr
 450 455 460
 Tyr Glu Asp Pro Asn Gln Ala Val Arg Glu Phe Ala Lys Glu Ile Asp
 465 470 475 480
 Ala Ser Cys Ile Lys Ile Glu Lys Val Ile Gly Val Gly Glu Phe Gly
 485 490 495
 Glu Val Cys Ser Gly Arg Leu Lys Val Pro Gly Lys Arg Glu Ile Cys
 500 505 510
 Val Ala Ile Lys Thr Leu Lys Ala Gly Tyr Thr Asp Lys Gln Arg Arg
 515 520 525
 Asp Phe Leu Ser Glu Ala Ser Ile Met Gly Gln Phe Asp His Pro Asn
 530 535 540
 Ile Ile His Leu Glu Gly Val Val Thr Lys Cys Lys Pro Val Met Ile
 545 550 555 560
 Ile Thr Glu Tyr Met Glu Asn Gly Ser Leu Asp Ala Phe Leu Arg Lys
 565 570 575
 Asn Asp Gly Arg Phe Thr Val Ile Gln Leu Val Gly Met Leu Arg Gly
 580 585 590
 Ile Gly Ser Gly Met Lys Tyr Leu Ser Asp Met Ser Tyr Val His Arg
 595 600 605
 Asp Leu Ala Ala Arg Asn Ile Leu Val Asn Ser Asn Leu Val Cys Lys
 610 615 620
 Val Ser Asp Phe Gly Met Ser Arg Val Leu Glu Asp Asp Pro Glu Ala
 625 630 635 640
 Ala Tyr Thr Thr Arg Gly Gly Lys Ile Pro Ile Arg Trp Thr Ala Pro
 645 650 655
 Glu Ala Ile Ala Tyr Arg Lys Phe Thr Ser Ala Ser Asp Val Trp Ser
 660 665 670
 Tyr Gly Ile Val Met Trp Glu Val Met Ser Tyr Gly Glu Arg Pro Tyr
 675 680 685
 Trp Asp Met Ser Asn Gln Asp Val Ile Lys Ala Ile Glu Glu Gly Tyr
 690 695 700
 Arg Leu Pro Pro Pro Met Asp Cys Pro Ile Ala Leu His Gln Leu Met
 705 710 715 720

56

Leu Asp Cys Trp Gln Lys Glu Arg Ser Asp Arg Pro Lys Phe Gly Gln
 725 730 735
 Ile Val Asn Met Leu Asp Lys Leu Ile Arg Asn Pro Asn Ser Leu Lys
 740 745 750
 Arg Thr Gly Ser Glu Ser Ser Arg Pro Ser Thr Ala Leu Leu Asp Pro
 755 760 765
 Ser Ser Pro Glu Phe Ser Ala Val Val Ser Val Ser Asp Trp Leu Gln
 770 775 780
 Ala Ile Lys Met Glu Arg Tyr Lys Asp Asn Phe Thr Ala Ala Gly Tyr
 785 790 795 800
 Thr Thr Leu Glu Ala Val Val His Met Asn Gln Asp Asp Leu Ala Arg
 805 810 815
 Ile Gly Ile Thr Ala Ile Thr His Gln Asn Lys Ile Leu Ser Ser Val
 820 825 830
 Gln Ala Met Arg Ser Gln Met Gln Gln Met His Gly Arg Met Val Pro
 835 840 845
 Val

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3776 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 290..3208

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGGCTCTGAC TTTGTGTTAA CGGTTTATGG ACTGGTTCCA AAGAGCTCAA AGGTACCAAA	60
ACACTCCAAG CAACCTCTGA ACCATTCAAG CAAGTAGTGT GTGTTTATTG GATATGGTGG	120
AGTCTACAGA GAATCTTCAT GGATTCTAAT GCTGACATCA GTGCAAGAAG AGTGTCAGGA	180
ATGGATTGGC TCTGGCTGGT TTGCTTCTTT CATCTAGTCA CTTCACTAGA AGACCTGCAT	240
CCTGACCAAC CGGAAAGGTG AGCAGGATGA GGCCATTGGT GGTGCTGTC ATG ACT	295
Met Thr	
1	
GAA ATA CTT CTG GAT ACA ACT GGA GAA ACC TCA GAG ATT GGC TGG ACC	343
Glu Ile Leu Leu Asp Thr Thr Gly Glu Thr Ser Glu Ile Gly Trp Thr	
5 10 15	
TCT CAC CCT CCT GAT GGG TGG GAA GAA GTA AGT GTC CGG GAT GAT AAG	391
Ser His Pro Pro Asp Gly Trp Glu Glu Val Ser Val Arg Asp Asp Lys	
20 25 30	
GAG CGC CAG ATC CGA ACC TTT CAA GTT TGT AAC ATG GAT GAA CCA GGT	439
Glu Arg Gln Ile Arg Thr Phe Gln Val Cys Asn Met Asp Glu Pro Gly	
35 40 45 50	

57

CAG AAT AAC TGG TTG CGT ACT CAC TTC ATA GAG CGA CGT GGA GCC CAC	487
Gln Asn Asn Trp Leu Arg Thr His Phe Ile Glu Arg Arg Gly Ala His	
55 60 65	
CGA GTC CAT GTC CGC CTT CAT TTC TCA GTG AGG GAC TGT GCC AGC ATG	535
Arg Val His Val Arg Leu His Phe Ser Val Arg Asp Cys Ala Ser Met	
70 75 80	
CGT ACT GTG GCC TCT ACT TGC AAA GAG ACT TTC ACA CTC TAC TAC CAC	583
Arg Thr Val Ala Ser Thr Cys Lys Glu Thr Phe Thr Leu Tyr Tyr His	
85 90 95	
CAG TCA GAT GTC GAC ATA GCC TCT CAG GAA CTG CCA GAG TGG CAT GAA	631
Gln Ser Asp Val Asp Ile Ala Ser Gln Glu Leu Pro Glu Trp His Glu	
100 105 110	
GGC CCC TGG ACC AAG GTG GAT ACT ATT GCA GCT GAT GAA AGC TTT TCC	679
Gly Pro Trp Thr Lys Val Asp Thr Ile Ala Ala Asp Glu Ser Phe Ser	
115 120 125 130	
CAG GTG GAC AGA ACT GGG AAG GTG GTA AGG ATG AAT GTT AAA GTA CGC	727
Gln Val Asp Arg Thr Gly Lys Val Val Arg Met Asn Val Lys Val Arg	
135 140 145	
AGC TTT GGG CCA CTC ACA AAG CAT GGC TTC TAC CTG GCC TTC CAG GAC	775
Ser Phe Gly Pro Leu Thr Lys His Gly Phe Tyr Leu Ala Phe Gln Asp	
150 155 160	
TCA GGA GCC TGT ATG TCC CTG GTG GCA GTC CAA GTC TTT TTC TAC AAG	823
Ser Gly Ala Cys Met Ser Leu Val Ala Val Gln Val Phe Phe Tyr Lys	
165 170 175	
TGT CCA GCT GTG GTG AAA GGA TTT GCC TCC TTC CCT GAA ACT TTT GCT	871
Cys Pro Ala Val Val Lys Gly Phe Ala Ser Phe Pro Glu Thr Phe Ala	
180 185 190	
GGA GGA GAG AGG ACC TCA CTG GTG GAG TCA CTA GGG ACG TGT GTA GCA	919
Gly Gly Glu Arg Thr Ser Leu Val Glu Ser Leu Gly Thr Cys Val Ala	
195 200 205 210	
AAT GCT GAA GAG GCA AGC ACA ACT GGG TCA TCA GGT GTT CGG TTG CAC	967
Asn Ala Glu Glu Ala Ser Thr Thr Gly Ser Ser Gly Val Arg Leu His	
215 220 225	
TGC AAT GGA GAA GGA GAG TGG ATG GTG GCC ACT GGA CGA TGC TCT TGC	1015
Cys Asn Gly Glu Gly Glu Trp Met Val Ala Thr Gly Arg Cys Ser Cys	
230 235 240	
AAG GCT GGT TAC CAA TCT GTT GAC AAT GAG CAA GCT TGT CAA GCT TGT	1063
Lys Ala Gly Tyr Gln Ser Val Asp Asn Glu Gln Ala Cys Gln Ala Cys	
245 250 255	
CCC ATT GGT TCC TTT AAA GCA TCT GTG GGA GAT GAC CCT TGC CTT CTC	1111
Pro Ile Gly Ser Phe Lys Ala Ser Val Gly Asp Asp Pro Cys Leu Leu	
260 265 270	
TGC CCT GCC CAC AGC CAT GCT CCA CTG CCA CTG CCA GGT TCC ATT GAA	1159
Cys Pro Ala His Ser His Ala Pro Leu Pro Leu Pro Gly Ser Ile Glu	
275 280 285 290	
TGT GTG TGT CAG AGT CAC TAC TAC CGA TCT GCT TCT GAC AAT TCT GAT	1207
Cys Val Cys Gln Ser His Tyr Tyr Arg Ser Ala Ser Asp Asn Ser Asp	
295 300 305	
GCT CCC TGC ACT GGC ATC CCC TCT GCT CCC CGT GAC CTC AGT TAT GAA	1255
Ala Pro Cys Thr Gly Ile Pro Ser Ala Pro Arg Asp Leu Ser Tyr Glu	
310 315 320	

ATT GTT GGC TCC AAC GTG CTC CTG ACC TGG CGC CTC CCC AAG GAC TTG Ile Val Gly Ser Asn Val Leu Leu Thr Trp Arg Leu Pro Lys Asp Leu 325 330 335	1303
GGT GGC CGC AAG GAT GTC TTC TTC AAT GTC ATC TGC AAG GAA TGC CCA Gly Gly Arg Lys Asp Val Phe Phe Asn Val Ile Cys Lys Glu Cys Pro 340 345 350	1351
ACA AGG TCA GCA GGG ACA TGT GTG CGC TGT GGG GAC AAT GTA CAG TTT Thr Arg Ser Ala Gly Thr Cys Val Arg Cys Gly Asp Asn Val Gln Phe 355 360 365 370	1399
GAA CCA CGC CAA GTG GGC CTG ACA GAA AGT CGT GTT CAA GTC TCC AAC Glu Pro Arg Gln Val Gly Leu Thr Glu Ser Arg Val Gln Val Ser Asn 375 380 385	1447
CTA TTG GCC CGT GTG CAG TAC ACT TTT GAG ATC CAG GCT GTC AAT TTG Leu Leu Ala Arg Val Gln Tyr Thr Phe Glu Ile Gln Ala Val Asn Leu 390 395 400	1495
GTG ACT GAG TTG AGT TCA GAA GCA CCC CAG TAT GCT ACC ATC AAC GTT Val Thr Glu Leu Ser Ser Glu Ala Pro Gln Tyr Ala Thr Ile Asn Val 405 410 415	1543
AGC ACC AGC CAG TCA GTG CCC TCC GCA ATC CCT ATG ATG CAT CAG GTG Ser Thr Ser Gln Ser Val Pro Ser Ala Ile Pro Met Met His Gln Val 420 425 430	1591
AGT CGT GCT ACC AGT AGC ATC ACA CTG TCT TGG CCT CAG CCA GAC CAG Ser Arg Ala Thr Ser Ser Ile Thr Leu Ser Trp Pro Gln Pro Asp Gln 435 440 445 450	1639
CCC AAT GGG GTT ATC CTG GAT TAC CAG CTA CGG TAC TTT GAC AAG GCA Pro Asn Gly Val Ile Leu Asp Tyr Gln Leu Arg Tyr Phe Asp Lys Ala 455 460 465	1687
GAA GAT GAG GAT AAT TCA TTT ACT TTG ACT AGT GAA ACT AAC ATG GCC Glu Asp Glu Asp Asn Ser Phe Thr Leu Thr Ser Glu Thr Asn Met Ala 470 475 480	1735
ACT ATA TTA AAT CTG AGT CCA GGC AAG ATC TAT GTC TTC CAA GTA CGA Thr Ile Leu Asn Leu Ser Pro Gly Lys Ile Tyr Val Phe Gln Val Arg 485 490 495	1783
GCT AGA ACA GCA GTG GGT TAT GGC CCA TAC AGT GGA AAG ATG TAT TTC Ala Arg Thr Ala Val Gly Tyr Gly Pro Tyr Ser Gly Lys Met Tyr Phe 500 505 510	1831
CAG ACT TTA ATG GCA GGA GAG CAC TCG GAG ATG GCA CAG GAC CGA CTG Gln Thr Leu Met Ala Gly Glu His Ser Glu Met Ala Gln Asp Arg Leu 515 520 525 530	1879
CCA CTT ATT GTG GGC TCA GCA CTT GGT GGT CTG GCA TTC TTG GTA ATT Pro Leu Ile Val Gly Ser Ala Leu Gly Gly Leu Ala Phe Leu Val Ile 535 540 545	1927
GCT GCC ATT GCC ATT CTT GCC ATC ATC TTC AAG AGT AAA AGG CGA GAG Ala Ala Ile Ala Ile Leu Ala Ile Ile Phe Lys Ser Lys Arg Arg Glu 550 555 560	1975
ACT CCA TAC ACA GAC CGC CTG CAG CAG TAT ATC AGT ACA CGA GGA CTT Thr Pro Tyr Thr Asp Arg Leu Gln Gln Tyr Ile Ser Thr Arg Gly Leu 565 570 575	2023
GGA GTG AAG TAT TAC ATT GAT CCT TCC ACG TAT GAA GAT CCC AAT GAA Gly Val Lys Tyr Tyr Ile Asp Pro Ser Thr Tyr Glu Asp Pro Asn Glu 580 585 590	2071

GCT ATT CGA GAG TTT GCC AAA GAG ATA GAT GTG TCC TTC ATC AAA ATT Ala Ile Arg Glu Phe Ala Lys Glu Ile Asp Val Ser Phe Ile Lys Ile 595 600 605 610	2119
GAG GAG GTC ATT GGA TCA GGA GAA TTT GGA GAG GTG TGC TTT GGG CGC Glu Glu Val Ile Gly Ser Gly Glu Phe Gly Glu Val Cys Phe Gly Arg 615 620 625	2167
CTA AAA CAC CCA GGG AAA CGT GAA TAC ACA GTA GCT ATT AAA ACC CTG Leu Lys His Pro Gly Lys Arg Glu Tyr Thr Val Ala Ile Lys Thr Leu 630 635 640	2215
AAG TCA GGT TAT ACT GAT GAA CAG CGT CGA GAG TTC CTG AGC GAG GCC Lys Ser Gly Tyr Thr Asp Glu Gln Arg Arg Glu Phe Leu Ser Glu Ala 645 650 655	2263
AGC ATC ATG GGG CAA TTT GAG CAT CCC AAT GTC ATC CAC CTG GAG GGC Ser Ile Met Gly Gln Phe Glu His Pro Asn Val Ile His Leu Glu Gly 660 665 670	2311
GTG GTC ACC AAA AGC CGA CCA GTC ATG ATT GTC ACA GAA TTC ATG GAG Val Val Thr Lys Ser Arg Pro Val Met Ile Val Thr Glu Phe Met Glu 675 680 685 690	2359
AAT GGA TCA CTG GAT TCC TTC CTC AGG GAG AAG GAG GGA CAG TTC AGT Asn Gly Ser Leu Asp Ser Phe Leu Arg Glu Lys Glu Gly Gln Phe Ser 695 700 705	2407
GTG TTA CAG CTG GTG GGA ATG CTA CGA GGG ATT GCA GCA GGC ATG CGC Val Leu Gln Leu Val Gly Met Leu Arg Gly Ile Ala Ala Gly Met Arg 710 715 720	2455
TAC CTT TCA GAC ATG AAC TAT GTG CAT CGT GAT CTC GCA GCA CGT AAC Tyr Leu Ser Asp Met Asn Tyr Val His Arg Asp Leu Ala Ala Arg Asn 725 730 735	2503
ATC TTA GTC AAC AGT AAC CTT GTA TGC AAG GTG TCA GAC TTT GGT TTG Ile Leu Val Asn Ser Asn Leu Val Cys Lys Val Ser Asp Phe Gly Leu 740 745 750	2551
TCT CGC TTT CTG GAA GAT GAT GCT TCA AAT CCC ACT TAT ACT GGA GCT Ser Arg Phe Leu Glu Asp Asp Ala Ser Asn Pro Thr Tyr Thr Gly Ala 755 760 765 770	2599
CTG GGT TGC AAA ATC CCC ATC CGT TGG ACT GCC CCT GAA GCT GTC CAG Leu Gly Cys Lys Ile Pro Ile Arg Trp Thr Ala Pro Glu Ala Val Gln 775 780 785	2647
TAT CGC AAG TTC ACC TCC TCC AGT GAT GTC TGG AGC TAT GGC ATT GTC Tyr Arg Lys Phe Thr Ser Ser Ser Asp Val Trp Ser Tyr Gly Ile Val 790 795 800	2695
ATG TGG GAG GTG ATG TCC TAT GGT GAG AGA CCT TAC TGG GAC ATG TCC Met Trp Glu Val Met Ser Tyr Gly Glu Arg Pro Tyr Trp Asp Met Ser 805 810 815	2743
AAC CAG GAT GTA ATT AAT GCC ATT GAC CAG GAC TAT CGC CTG CCA CCA Asn Gln Asp Val Ile Asn Ala Ile Asp Gln Asp Tyr Arg Leu Pro Pro 820 825 830	2791
CCC CCA GAC TGC CCA ACT GTT TTG CAT CTG CTG ATG CTT GAC TGC TGG Pro Pro Asp Cys Pro Thr Val Leu His Leu Leu Met Leu Asp Cys Trp 835 840 845 850	2839
CAG AAG GAT CGA GTC CAG AGA CCA AAA TTT GAA CAA ATA GTC AGT GCC Gln Lys Asp Arg Val Gln Arg Pro Lys Phe Glu Gln Ile Val Ser Ala 855 860 865	2887

60

CTA GAT AAA ATG ATC CGC AAG CCA TCT GCT CTC AAA GCC ACT GGC ACT Leu Asp Lys Met Ile Arg Lys Pro Ser Ala Leu Lys Ala Thr Gly Thr 870 875 880	2935
GGG AGC AGC AGA CCA TCT CAG CCT CTC CTG AGC AAC TCC CCT CCA GAT Gly Ser Ser Arg Pro Ser Gln Pro Leu Leu Ser Asn Ser Pro Pro Asp 885 890 895	2983
TTT CCT TCA CTC AGC AAT GCC CAC GAG TGG TTG GAT GCC ATC AAG ATG Phe Pro Ser Leu Ser Asn Ala His Glu Trp Leu Asp Ala Ile Lys Met 900 905 910	3031
GGT CGT TAC AAG GAG AAT TTT GAC CAG GCT GGT CTG ATT ACA TTT GAT Gly Arg Tyr Lys Glu Asn Phe Asp Gln Ala Gly Leu Ile Thr Phe Asp 915 920 925 930	3079
GTC ATA TCA CGC ATG ACT CTG GAA GAT CTC CAG CGT ATT GGA ATC ACC Val Ile Ser Arg Met Thr Leu Glu Asp Leu Gln Arg Ile Gly Ile Thr 935 940 945	3127
CTG GTT GGT CAC CAG AAA AAG ATT CTA AAC AGC ATC CAG CTC ATG AAA Leu Val Gly His Gln Lys Lys Ile Leu Asn Ser Ile Gln Leu Met Lys 950 955 960	3175
GTT CAT TTG AAC CAG CTT GAA CCA GTT GAA GTG TGATGCTTTA AGTCTCTATT Val His Leu Asn Gln Leu Glu Pro Val Glu Val 965 970	3228
TCACCAGACT CAAATTCTGA AAGAGTCCTG AGGGGATTCA GAGGGATTGT CACTGTATGA	3288
AAAGGAAATG GCAAGATGCT CCTTGAAGAC TTA CTGCACC TAGAGAGTAG ACATTACACA	3348
TTCCATTCCA CCAGCAAAAA GAGAATCTTG CCATCATTTA AAAGCAGAGT TAAATAGCTG	3408
GTGGTTAAAT ATGACTGGCA TCATACACTA GGAGTAGGTC AGGGAGGGAA AGTTATAGTA	3468
ATGCAGAGTG GAGCTGGTAT AATAGTTTGG ACAGACCACA AGCACCTGCT AGCTCTTCTC	3528
CACTAAATAA AAAATCAGAC AATTCTCCAG TGCCATCAGC AGGCTTTATC TGTGACTGGG	3588
AACAAAGAAA TCACAATTTT TCCAAGAGAG TATCAGCACA TTGTGAGAGT TATCACTCAG	3648
TTGGAAATGG ACATCACTTG CTATGCCAGA TTTGTGAGAA ACTGGAGTTC CACTGAGTGC	3708
ACCATATGTG GTAAACAATA AGGTACATCA CCTCGTAATT TTTACAGAGG TTGAGAGTAA	3768
AGGGCCCA	3776

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 973 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Thr Glu Ile Leu Leu Asp Thr Thr Gly Glu Thr Ser Glu Ile Gly
 1 5 10 15
 Trp Thr Ser His Pro Pro Asp Gly Trp Glu Glu Val Ser Val Arg Asp
 20 25 30

61

Asp Lys Glu Arg Gln Ile Arg Thr Phe Gln Val Cys Asn Met Asp Glu
 35 40 45
 Pro Gly Gln Asn Asn Trp Leu Arg Thr His Phe Ile Glu Arg Arg Gly
 50 55 60
 Ala His Arg Val His Val Arg Leu His Phe Ser Val Arg Asp Cys Ala
 65 70 75 80
 Ser Met Arg Thr Val Ala Ser Thr Cys Lys Glu Thr Phe Thr Leu Tyr
 85 90 95
 Tyr His Gln Ser Asp Val Asp Ile Ala Ser Gln Glu Leu Pro Glu Trp
 100 105 110
 His Glu Gly Pro Trp Thr Lys Val Asp Thr Ile Ala Ala Asp Glu Ser
 115 120 125
 Phe Ser Gln Val Asp Arg Thr Gly Lys Val Val Arg Met Asn Val Lys
 130 135 140
 Val Arg Ser Phe Gly Pro Leu Thr Lys His Gly Phe Tyr Leu Ala Phe
 145 150 155 160
 Gln Asp Ser Gly Ala Cys Met Ser Leu Val Ala Val Gln Val Phe Phe
 165 170 175
 Tyr Lys Cys Pro Ala Val Val Lys Gly Phe Ala Ser Phe Pro Glu Thr
 180 185 190
 Phe Ala Gly Gly Glu Arg Thr Ser Leu Val Glu Ser Leu Gly Thr Cys
 195 200 205
 Val Ala Asn Ala Glu Glu Ala Ser Thr Thr Gly Ser Ser Gly Val Arg
 210 215 220
 Leu His Cys Asn Gly Glu Gly Glu Trp Met Val Ala Thr Gly Arg Cys
 225 230 235 240
 Ser Cys Lys Ala Gly Tyr Gln Ser Val Asp Asn Glu Gln Ala Cys Gln
 245 250 255
 Ala Cys Pro Ile Gly Ser Phe Lys Ala Ser Val Gly Asp Asp Pro Cys
 260 265 270
 Leu Leu Cys Pro Ala His Ser His Ala Pro Leu Pro Leu Pro Gly Ser
 275 280 285
 Ile Glu Cys Val Cys Gln Ser His Tyr Tyr Arg Ser Ala Ser Asp Asn
 290 295 300
 Ser Asp Ala Pro Cys Thr Gly Ile Pro Ser Ala Pro Arg Asp Leu Ser
 305 310 315 320
 Tyr Glu Ile Val Gly Ser Asn Val Leu Leu Thr Trp Arg Leu Pro Lys
 325 330 335
 Asp Leu Gly Gly Arg Lys Asp Val Phe Phe Asn Val Ile Cys Lys Glu
 340 345 350
 Cys Pro Thr Arg Ser Ala Gly Thr Cys Val Arg Cys Gly Asp Asn Val
 355 360 365
 Gln Phe Glu Pro Arg Gln Val Gly Leu Thr Glu Ser Arg Val Gln Val
 370 375 380

62

Ser Asn Leu Leu Ala Arg Val Gln Tyr Thr Phe Glu Ile Gln Ala Val
 385 390 395 400
 Asn Leu Val Thr Glu Leu Ser Ser Glu Ala Pro Gln Tyr Ala Thr Ile
 405 410 415
 Asn Val Ser Thr Ser Gln Ser Val Pro Ser Ala Ile Pro Met Met His
 420 425 430
 Gln Val Ser Arg Ala Thr Ser Ser Ile Thr Leu Ser Trp Pro Gln Pro
 435 440 445
 Asp Gln Pro Asn Gly Val Ile Leu Asp Tyr Gln Leu Arg Tyr Phe Asp
 450 455 460
 Lys Ala Glu Asp Glu Asp Asn Ser Phe Thr Leu Thr Ser Glu Thr Asn
 465 470 475 480
 Met Ala Thr Ile Leu Asn Leu Ser Pro Gly Lys Ile Tyr Val Phe Gln
 485 490 495
 Val Arg Ala Arg Thr Ala Val Gly Tyr Gly Pro Tyr Ser Gly Lys Met
 500 505 510
 Tyr Phe Gln Thr Leu Met Ala Gly Glu His Ser Glu Met Ala Gln Asp
 515 520 525
 Arg Leu Pro Leu Ile Val Gly Ser Ala Leu Gly Gly Leu Ala Phe Leu
 530 535 540
 Val Ile Ala Ala Ile Ala Ile Leu Ala Ile Ile Phe Lys Ser Lys Arg
 545 550 555 560
 Arg Glu Thr Pro Tyr Thr Asp Arg Leu Gln Gln Tyr Ile Ser Thr Arg
 565 570 575
 Gly Leu Gly Val Lys Tyr Tyr Ile Asp Pro Ser Thr Tyr Glu Asp Pro
 580 585 590
 Asn Glu Ala Ile Arg Glu Phe Ala Lys Glu Ile Asp Val Ser Phe Ile
 595 600 605
 Lys Ile Glu Glu Val Ile Gly Ser Gly Glu Phe Gly Glu Val Cys Phe
 610 615 620
 Gly Arg Leu Lys His Pro Gly Lys Arg Glu Tyr Thr Val Ala Ile Lys
 625 630 635 640
 Thr Leu Lys Ser Gly Tyr Thr Asp Glu Gln Arg Arg Glu Phe Leu Ser
 645 650 655
 Glu Ala Ser Ile Met Gly Gln Phe Glu His Pro Asn Val Ile His Leu
 660 665 670
 Glu Gly Val Val Thr Lys Ser Arg Pro Val Met Ile Val Thr Glu Phe
 675 680 685
 Met Glu Asn Gly Ser Leu Asp Ser Phe Leu Arg Glu Lys Glu Gly Gln
 690 695 700
 Phe Ser Val Leu Gln Leu Val Gly Met Leu Arg Gly Ile Ala Ala Gly
 705 710 715 720
 Met Arg Tyr Leu Ser Asp Met Asn Tyr Val His Arg Asp Leu Ala Ala
 725 730 735

63

Arg Asn Ile Leu Val Asn Ser Asn Leu Val Cys Lys Val Ser Asp Phe
 740 745 750
 Gly Leu Ser Arg Phe Leu Glu Asp Asp Ala Ser Asn Pro Thr Tyr Thr
 755 760 765
 Gly Ala Leu Gly Cys Lys Ile Pro Ile Arg Trp Thr Ala Pro Glu Ala
 770 775 780
 Val Gln Tyr Arg Lys Phe Thr Ser Ser Ser Asp Val Trp Ser Tyr Gly
 785 790 795 800
 Ile Val Met Trp Glu Val Met Ser Tyr Gly Glu Arg Pro Tyr Trp Asp
 805 810 815
 Met Ser Asn Gln Asp Val Ile Asn Ala Ile Asp Gln Asp Tyr Arg Leu
 820 825 830
 Pro Pro Pro Pro Asp Cys Pro Thr Val Leu His Leu Leu Met Leu Asp
 835 840 845
 Cys Trp Gln Lys Asp Arg Val Gln Arg Pro Lys Phe Glu Gln Ile Val
 850 855 860
 Ser Ala Leu Asp Lys Met Ile Arg Lys Pro Ser Ala Leu Lys Ala Thr
 865 870 875 880
 Gly Thr Gly Ser Ser Arg Pro Ser Gln Pro Leu Leu Ser Asn Ser Pro
 885 890 895
 Pro Asp Phe Pro Ser Leu Ser Asn Ala His Glu Trp Leu Asp Ala Ile
 900 905 910
 Lys Met Gly Arg Tyr Lys Glu Asn Phe Asp Gln Ala Gly Leu Ile Thr
 915 920 925
 Phe Asp Val Ile Ser Arg Met Thr Leu Glu Asp Leu Gln Arg Ile Gly
 930 935 940
 Ile Thr Leu Val Gly His Gln Lys Lys Ile Leu Asn Ser Ile Gln Leu
 945 950 955 960
 Met Lys Val His Leu Asn Gln Leu Glu Pro Val Glu Val
 965 970

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3546 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..2920

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

C GGG GTC TCC TCG AGG GCG CGG CCG CCG GGC AGC AGC AGG AGC
 Gly Val Ser Ser Arg Ala Arg Pro Pro Gly Ser Ser Arg Ser
 1 5 10 15

AGC AGG AGG GGG GTG ACC TCG GAG CTG GCA TGG ACA ACC CAT CCG GAG Ser Arg Arg Gly Val Thr Ser Glu Leu Ala Trp Thr Thr His Pro Glu	94
20 25 30	
ACG GGG TGG GAA GAG GTC AGT GGT TAC GAC GAG GCT ATG AAC CCC ATC Thr Gly Trp Glu Glu Val Ser Gly Tyr Asp Glu Ala Met Asn Pro Ile	142
35 40 45	
CGC ACA TAC CAG GTG TGC AAC GTG CGG GAG GCC AAC CAG AAC AAC TGG Arg Thr Tyr Gln Val Cys Asn Val Arg Glu Ala Asn Gln Asn Asn Trp	190
50 55 60	
CTT CGC ACC AAG TTC ATT CAG CGC CAG GAC GTC CAG CGT GTC TAC GTG Leu Arg Thr Lys Phe Ile Gln Arg Gln Asp Val Gln Arg Val Tyr Val	238
65 70 75	
GAG CTG AAA TTC ACT GTG CGG GAC TGC AAC AGC ATC CCC AAC ATC CCT Glu Leu Lys Phe Thr Val Arg Asp Cys Asn Ser Ile Pro Asn Ile Pro	286
80 85 90 95	
GGT TCC TGC AAA GAG ACC TTC AAC CTC TTC TAT TAT GAG TCA GAT ACG Gly Ser Cys Lys Glu Thr Phe Asn Leu Phe Tyr Tyr Glu Ser Asp Thr	334
100 105 110	
GAT TCT GCC TCT GCC AAT AGC CCT TTC TGG ATG GAG AAC CCC TAT ATC Asp Ser Ala Ser Ala Asn Ser Pro Phe Trp Met Glu Asn Pro Tyr Ile	382
115 120 125	
AAA GTG GAT ACA ATT GCT CCG GAT GAG AGC TTC TCC AAA CTG GAG TCC Lys Val Asp Thr Ile Ala Pro Asp Glu Ser Phe Ser Lys Leu Glu Ser	430
130 135 140	
GGC CGT GTG AAC ACC AAG GTG CGC AGC TTT GGG CCG CTC TCC AAG AAT Gly Arg Val Asn Thr Lys Val Arg Ser Phe Gly Pro Leu Ser Lys Asn	478
145 150 155	
GGC TTT TAT CTG GCT TTC CAG GAC CTG GGG GCC TGC ATG TCC CTT ATC Gly Phe Tyr Leu Ala Phe Gln Asp Leu Gly Ala Cys Met Ser Leu Ile	526
160 165 170 175	
TCC GTC CGG GCT TTC TAC AAG AAA TGT TCC AAC ACC ATC GCT GGC TTT Ser Val Arg Ala Phe Tyr Lys Lys Cys Ser Asn Thr Ile Ala Gly Phe	574
180 185 190	
GCT ATC TTC CCG GAG ACC CTA ACG GGG GCT GAG CCC ACG TCG CTG GTC Ala Ile Phe Pro Glu Thr Leu Thr Gly Ala Glu Pro Thr Ser Leu Val	622
195 200 205	
ATT GCG CCG GGC ACC TGC ATC CCC AAC GCA GTG GAA GTG TCT GTG CCC Ile Ala Pro Gly Thr Cys Ile Pro Asn Ala Val Glu Val Ser Val Pro	670
210 215 220	
CTG AAG CTG TAC TGC AAC GGT GAT GGC GAG TGG ATG GTG CCT GTG GGA Leu Lys Leu Tyr Cys Asn Gly Asp Gly Glu Trp Met Val Pro Val Gly	718
225 230 235	
GCG TGC ACG TGT GCT GCT GGG TAC GAG CCA GCC ATG AAG GAT ACC CAG Ala Cys Thr Cys Ala Ala Gly Tyr Glu Pro Ala Met Lys Asp Thr Gln	766
240 245 250 255	
TGC CAA GCA TGC GGC CCG GGG ACG TTC AAA TCC AAG CAG GGC GAG GGC Cys Gln Ala Cys Gly Pro Gly Thr Phe Lys Ser Lys Gln Gly Glu Gly	814
260 265 270	
CCC TGC TCC CCC TGC CCT CCC AAC AGC CGC ACC ACC GCG GGG GCA GCC Pro Cys Ser Pro Cys Pro Pro Asn Arg Thr Thr Ala Gly Ala Ala	862
275 280 285	

65

ACA GTC TGC ATA TGT CGC AGC GGC TTC TTC CGA GCA GAC GCG GAC CCC Thr Val Cys Ile Cys Arg Ser Gly Phe Phe Arg Ala Asp Ala Asp Pro 290 295 300	910
GCA GAC AGC GCC TGC ACC AGT GTG CCC TCA GCC CCA CGC AGC GTC ATC Ala Asp Ser Ala Cys Thr Ser Val Pro Ser Ala Pro Arg Ser Val Ile 305 310 315	958
TCC AAC GTG AAT GAG ACG TCG TTG GTG CTG GAG TGG AGC GAG CCG CAG Ser Asn Val Asn Glu Thr Ser Leu Val Leu Glu Trp Ser Glu Pro Gln 320 325 330 335	1006
GAC GCG GGC GGG CGG GAT GAC CTG CTC TAC AAC GTC ATC TGC AAG AAG Asp Ala Gly Gly Arg Asp Asp Leu Leu Tyr Asn Val Ile Cys Lys Lys 340 345 350	1054
TGC AGC GTG GAG CGG CGG CTG TGC AGC CGC TGC GAC GAC AAC GTG GAG Cys Ser Val Glu Arg Arg Leu Cys Ser Arg Cys Asp Asp Asn Val Glu 355 360 365	1102
TTC GTG CCG CGC CAG CTG GGC CTC ACT GGC CTC ACT GAG CGA CGC ATC Phe Val Pro Arg Gln Leu Gly Thr Thr Gly Leu Thr Glu Arg Arg Ile 370 375 380	1150
TAC ATC AGC AAG GTG ATG GCC CAC CCC CAG TAC ACC TTC GAG ATC CAG Tyr Ile Ser Lys Val Met Ala His Pro Gln Tyr Thr Phe Glu Ile Gln 385 390 395	1198
GCG GTG AAT GGC ATC TCC AGC AAG AGC CCC TAC CCT CCC CAT TTT GCC Ala Val Asn Gly Ile Ser Ser Lys Ser Pro Tyr Pro Pro His Phe Ala 400 405 410 415	1246
TCC GTC AAC ATC ACG ACC AAC CAG GCA GCC CCA TCT GCC GTG CCC ACC Ser Val Asn Ile Thr Thr Asn Gln Ala Ala Pro Ser Ala Val Pro Thr 420 425 430	1294
ATG CAT CTG CAC AGC AGC ACC GGG AAC AGC ATG ACA CTG TCA TGG ACT Met His Leu His Ser Ser Thr Gly Asn Ser Met Thr Leu Ser Trp Thr 435 440 445	1342
CCC CCG GAA AGG CCC AAC GGC ATC ATT CTC GAC TAT GAA ATC AAG TAC Pro Pro Glu Arg Pro Asn Gly Ile Ile Leu Asp Tyr Glu Ile Lys Tyr 450 455 460	1390
TCC GAG AAG CAA GGC CAG GGT GAC GGC ATT GCC AAC ACT GTC ACC AGC Ser Glu Lys Gln Gly Gln Gly Asp Gly Ile Ala Asn Thr Val Thr Ser 465 470 475	1438
CAG AAG AAC TCG GTG CGG CTG GAC GGA CTG AAG GCC AAT GCT CGG TAC Gln Lys Asn Ser Val Arg Leu Asp Gly Leu Lys Ala Asn Ala Arg Tyr 480 485 490 495	1486
ATG GTG CAG GTC CGG GCG CGC ACA GTG GCT GGA TAC GGC CGC TAC AGC Met Val Gln Val Arg Ala Arg Thr Val Ala Gly Tyr Gly Arg Tyr Ser 500 505 510	1534
CTC CCC ACC GAG TTC CAG ACG ACT GCG GAG GAT GGC TCC ACC AGC AAG Leu Pro Thr Glu Phe Gln Thr Thr Ala Glu Asp Gly Ser Thr Ser Lys 515 520 525	1582
ACT TTC CAG GAG CTT CCT CTC ATC GTG GGT TCA GCC ACC GCG GGA CTG Thr Phe Gln Glu Leu Pro Leu Ile Val Gly Ser Ala Thr Ala Gly Leu 530 535 540	1630
CTG TTT GTC ATC GTG GTG GTC ATC ATC GCT ATT GTC TGC TTC AGG AAG Leu Phe Val Ile Val Val Val Ile Ile Ala Ile Val Cys Phe Arg Lys 545 550 555	1678

CAG CGC AAC AGC ACA GAT CCC GAG TAC ACA GAG AAG CTG CAG CAA TAT Gln Arg Asn Ser Thr Asp Pro Glu Tyr Thr Glu Lys Leu Gln Gln Tyr 560 565 570 575	1726
GTC ACT CCT GGG ATG AAG GTC TAC ATT GAC CCC TTC ACC TAT GAA GAC Val Thr Pro Gly Met Lys Val Tyr Ile Asp Pro Phe Thr Tyr Glu Asp 580 585 590	1774
CCA AAT GAA GCT GTC CGG GAA TTC GCC AAA GAG ATT GAT ATC TCC TGT Pro Asn Glu Ala Val Arg Glu Phe Ala Lys Glu Ile Asp Ile Ser Cys 595 600 605	1822
GTC AAA ATT GAG GAG GTC ATT GGA GCA GGA GAG TTT GGT GAG GTG TGC Val Lys Ile Glu Glu Val Ile Gly Ala Gly Glu Phe Gly Glu Val Cys 610 615 620	1870
CGT GGG CGC CTG AAG CTG CCT GGC CGC CGT GAG ATC TTT GTG GCC ATC Arg Gly Arg Leu Lys Leu Pro Gly Arg Arg Glu Ile Phe Val Ala Ile 625 630 635	1918
AAG ACA CTG AAG GTG GGC TAC ACA GAG AGG CAG CGG CGG GAC TTC CTG Lys Thr Leu Lys Val Gly Tyr Thr Glu Arg Gln Arg Arg Asp Phe Leu 640 645 650 655	1966
AGT GAG GCC AGC ATC ATG GGC CAG TTC GAC CAC CCC AAC ATC ATC CAC Ser Glu Ala Ser Ile Met Gly Gln Phe Asp His Pro Asn Ile Ile His 660 665 670	2014
CTG GAG GGC GTG GTG ACC AAG AGC CGC CCT GTC ATG ATC ATC ACA GAG Leu Glu Gly Val Val Thr Lys Ser Arg Pro Val Met Ile Ile Thr Glu 675 680 685	2062
TTC ATG GAG AAC TGC GCT CTC GAC TCC TTC CTC CGG CTG AAT GAT GGG Phe Met Glu Asn Cys Ala Leu Asp Ser Phe Leu Arg Leu Asn Asp Gly 690 695 700	2110
CAG TTC ACG GTC ATC CAG CTG GTG GGG ATG CTG CGA GGC ATC GCT GCT Gln Phe Thr Val Ile Gln Leu Val Gly Met Leu Arg Gly Ile Ala Ala 705 710 715	2158
GGC ATG AAG TAC CTC TCA GAG ATG AAC TAC GTG CAC CGA GAC CTG GCT Gly Met Lys Tyr Leu Ser Glu Met Asn Tyr Val His Arg Asp Leu Ala 720 725 730 735	2206
GCC CGC AAC ATC CTG GTC AAC AGC AAC TTG GTC TGC AAA GTG TCT GAC Ala Arg Asn Ile Leu Val Asn Ser Asn Leu Val Cys Lys Val Ser Asp 740 745 750	2254
TTC GGG CTC TCC CGC TTT TTG GAG GAT GAT CCA GCC GAC CCC ACC TAC Phe Gly Leu Ser Arg Phe Leu Glu Asp Asp Pro Ala Asp Pro Thr Tyr 755 760 765	2302
ACC AGC TCC CTG GGA GGC AAG ATC CCC ATC AGG TGG ACA GCT CCT GAG Thr Ser Ser Leu Gly Gly Lys Ile Pro Ile Arg Trp Thr Ala Pro Glu 770 775 780	2350
GCC ATC GCC TAC CGC AAA TTC ACG TCG GCC AGC GAC GTG TGG AGC TAC Ala Ile Ala Tyr Arg Lys Phe Thr Ser Ala Ser Asp Val Trp Ser Tyr 785 790 795	2398
GGC ATC GTC ATG TGG GAA GTG ATG TCC TAC GGG GAG CGA CCC TAC TGG Gly Ile Val Met Trp Glu Val Met Ser Tyr Glu Arg Pro Tyr Trp 800 805 810 815	2446
GAC ATG TCC AAC CAG GAT GTG ATC AAC GCG GTG GAG CAG GAT TAC CGC Asp Met Ser Asn Gln Asp Val Ile Asn Ala Val Glu Gln Asp Tyr Arg 820 825 830	2494

67

CTG CCA CCC CCC ATG GAC TGC CCC ACA GCA CTG CAC CAG CTG ATG CTG Leu Pro Pro Pro Met Asp Cys Pro Thr Ala Leu His Gln Leu Met Leu 835 840 845	2542
GAC TGC TGG GTG CGG GAC CGC AAC CTG CGG CCC AAG TTT GCA CAG ATT Asp Cys Trp Val Arg Asp Arg Asn Leu Arg Pro Lys Phe Ala Gln Ile 850 855 860	2590
GTC AAC ACG CTG GAC AAG CTG ATC CGC AAT GCT GCC AGC CTG AAG GTC Val Asn Thr Leu Asp Lys Ser Gly Val Ile Arg Asn Ala Ala Ser Leu Lys Val 865 870 875	2638
ATC GCC AGC GTC CAG TCC GGT GTC TCC CAG CCG CTC CTG GAC CGC ACC Ile Ala Ser Val Gln Ser Gly Val Ser Gln Pro Leu Leu Asp Arg Thr 880 885 890 895	2686
GTG CCC GAT TAC ACC ACC TTC ACC ACC GTG GGA GAC TGG CTG GAT GCC Val Pro Asp Tyr Thr Thr Phe Thr Thr Val Gly Asp Trp Leu Asp Ala 900 905 910	2734
ATC AAA ATG GGA CGG TAC AAG GAG AAC TTC GTC AAC GCC GGC TTC GCC Ile Lys Met Gly Arg Tyr Lys Glu Asn Phe Val Asn Ala Gly Phe Ala 915 920 925	2782
TCC TTT GAC CTG GTG GCA CAG ATG ACA GCA GAG GAC CTG CTA AGG ATA Ser Phe Asp Leu Val Ala Gln Met Thr Ala Glu Asp Leu Leu Arg Ile 930 935 940	2830
GGA GTG ACG CTA GCA GGG CAC CAG AAG AAG ATC CTG AGC AGC ATT CAG Gly Val Thr Leu Ala Gly His Gln Lys Lys Ile Leu Ser Ser Ile Gln 945 950 955	2878
GAC ATG AGG CTG CAG ATG AAC CAG ACG CTG CCG GTT CAG GTT Asp Met Arg Leu Gln Met Asn Gln Thr Leu Pro Val Gln Val 960 965 970	2920
TGACCGCAGG GACTCTGCAT TGGAACGGAC TGAGGGAACC TGCCAACCAG GTTCTGTTTG	2980
CGGTGCAGCC CGGCTTCCCG ATTTCCCTT CCCGTGGCGC TCCTCTGCCT CGGACGCTCG	3040
CCGGGGACAG GCTGGGCCCG GCCACCCTTC CCTGGATCAG AGGCACTCGT GCCGGGAGGG	3100
AGCCCGGGCTT TTCGTCCCGT GTCCCGCAGC GGCGAGGCAG TGAACGCAGT CTTCATATTG	3160
AAGATGGATT ATGGGACGGA GATGGCGCAT CCGCTTCCCG CCCTGTCTCA GTGCTCATCA	3220
GTITGAAGAG ATGTTCTGCT TCTTGGATT CTTTACACCC CGGTTTTTCCC CCCTCGAGTC	3280
CTCACTTCCC CCTATCCCTG AGGCCACAGA CTGTTGACCC GTCCGCTGAG TCCGTCAGAC	3340
GCTCCGAAGC CTTCCCGGAG CCCGGTCCCC GCGTGGAGAC GGCGCCAGGG ACGGGGCTAC	3400
GGCCCCAGAC AATCACTCCA CCCCTCCGCA CGAGGGTCCT CACTGGGACG TGTCTGAAGG	3460
GGAAAGGCTC TGCTCCCTTT TTGGCTTTGC ACGCCAGAAC CCGAACCCCG TGAGATTAC	3520
TATGCAGGGA GTTAGGCAAA AAAAAG	3546

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 973 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Gly Val Ser Ser Arg Ala Arg Arg Pro Pro Gly Ser Ser Arg Ser Ser
 1 5 10 15
 Arg Arg Gly Val Thr Ser Glu Leu Ala Trp Thr Thr His Pro Glu Thr
 20 25 30
 Gly Trp Glu Glu Val Ser Gly Tyr Asp Glu Ala Met Asn Pro Ile Arg
 35 40 45
 Thr Tyr Gln Val Cys Asn Val Arg Glu Ala Asn Gln Asn Asn Trp Leu
 50 55 60
 Arg Thr Lys Phe Ile Gln Arg Gln Asp Val Gln Arg Val Tyr Val Glu
 65 70 75 80
 Leu Lys Phe Thr Val Arg Asp Cys Asn Ser Ile Pro Asn Ile Pro Gly
 85 90 95
 Ser Cys Lys Glu Thr Phe Asn Leu Phe Tyr Tyr Glu Ser Asp Thr Asp
 100 105 110
 Ser Ala Ser Ala Asn Ser Pro Phe Trp Met Glu Asn Pro Tyr Ile Lys
 115 120 125
 Val Asp Thr Ile Ala Pro Asp Glu Ser Phe Ser Lys Leu Glu Ser Gly
 130 135 140
 Arg Val Asn Thr Lys Val Arg Ser Phe Gly Pro Leu Ser Lys Asn Gly
 145 150 155 160
 Phe Tyr Leu Ala Phe Gln Asp Leu Gly Ala Cys Met Ser Leu Ile Ser
 165 170 175
 Val Arg Ala Phe Tyr Lys Lys Cys Ser Asn Thr Ile Ala Gly Phe Ala
 180 185 190
 Ile Phe Pro Glu Thr Leu Thr Gly Ala Glu Pro Thr Ser Leu Val Ile
 195 200 205
 Ala Pro Gly Thr Cys Ile Pro Asn Ala Val Glu Val Ser Val Pro Leu
 210 215 220
 Lys Leu Tyr Cys Asn Gly Asp Gly Glu Trp Met Val Pro Val Gly Ala
 225 230 235 240
 Cys Thr Cys Ala Ala Gly Tyr Glu Pro Ala Met Lys Asp Thr Gln Cys
 245 250 255
 Gln Ala Cys Gly Pro Gly Thr Phe Lys Ser Lys Gln Gly Glu Gly Pro
 260 265 270
 Cys Ser Pro Cys Pro Pro Asn Ser Arg Thr Thr Ala Gly Ala Ala Thr
 275 280 285
 Val Cys Ile Cys Arg Ser Gly Phe Phe Arg Ala Asp Ala Asp Pro Ala
 290 295 300
 Asp Ser Ala Cys Thr Ser Val Pro Ser Ala Pro Arg Ser Val Ile Ser
 305 310 315 320
 Asn Val Asn Glu Thr Ser Leu Val Leu Glu Trp Ser Glu Pro Gln Asp
 325 330 335
 Ala Gly Gly Arg Asp Asp Leu Leu Tyr Asn Val Ile Cys Lys Lys Cys
 340 345 350

69

Ser Val Glu Arg Arg Leu Cys Ser Arg Cys Asp Asp Asn Val Glu Phe
 355 360 365
 Val Pro Arg Gln Leu Gly Leu Thr Gly Leu Thr Glu Arg Arg Ile Tyr
 370 375 380
 Ile Ser Lys Val Met Ala His Pro Gln Tyr Thr Phe Glu Ile Gln Ala
 385 390 395 400
 Val Asn Gly Ile Ser Ser Lys Ser Pro Tyr Pro Pro His Phe Ala Ser
 405 410 415
 Val Asn Ile Thr Thr Asn Gln Ala Ala Pro Ser Ala Val Pro Thr Met
 420 425 430
 His Leu His Ser Ser Thr Gly Asn Ser Met Thr Leu Ser Trp Thr Pro
 435 440 445
 Pro Glu Arg Pro Asn Gly Ile Ile Leu Asp Tyr Glu Ile Lys Tyr Ser
 450 455 460
 Glu Lys Gln Gly Gln Gly Asp Gly Ile Ala Asn Thr Val Thr Ser Gln
 465 470 475 480
 Lys Asn Ser Val Arg Leu Asp Gly Leu Lys Ala Asn Ala Arg Tyr Met
 485 490 495
 Val Gln Val Arg Ala Arg Thr Val Ala Gly Tyr Gly Arg Tyr Ser Leu
 500 505 510
 Pro Thr Glu Phe Gln Thr Thr Ala Glu Asp Gly Ser Thr Ser Lys Thr
 515 520 525
 Phe Gln Glu Leu Pro Leu Ile Val Gly Ser Ala Thr Ala Gly Leu Leu
 530 535 540
 Phe Val Ile Val Val Val Ile Ile Ala Ile Val Cys Phe Arg Lys Gln
 545 550 555 560
 Arg Asn Ser Thr Asp Pro Glu Tyr Thr Glu Lys Leu Gln Gln Tyr Val
 565 570 575
 Thr Pro Gly Met Lys Val Tyr Ile Asp Pro Phe Thr Tyr Glu Asp Pro
 580 585 590
 Asn Glu Ala Val Arg Glu Phe Ala Lys Glu Ile Asp Ile Ser Cys Val
 595 600 605
 Lys Ile Glu Glu Val Ile Gly Ala Gly Glu Phe Gly Glu Val Cys Arg
 610 615 620
 Gly Arg Leu Lys Leu Pro Gly Arg Arg Glu Ile Phe Val Ala Ile Lys
 625 630 635 640
 Thr Leu Lys Val Gly Tyr Thr Glu Arg Gln Arg Arg Asp Phe Leu Ser
 645 650 655
 Glu Ala Ser Ile Met Gly Gln Phe Asp His Pro Asn Ile Ile His Leu
 660 665 670
 Glu Gly Val Val Thr Lys Ser Arg Pro Val Met Ile Ile Thr Glu Phe
 675 680 685
 Met Glu Asn Cys Ala Leu Asp Ser Phe Leu Arg Leu Asn Asp Gly Gln
 690 695 700

70

Phe Thr Val Ile Gln Leu Val Gly Met Leu Arg Gly Ile Ala Ala Gly
 705 710 715 720
 Met Lys Tyr Leu Ser Glu Met Asn Tyr Val His Arg Asp Leu Ala Ala
 725 730 735
 Arg Asn Ile Leu Val Asn Ser Asn Leu Val Cys Lys Val Ser Asp Phe
 740 745 750
 Gly Leu Ser Arg Phe Leu Glu Asp Asp Pro Ala Asp Pro Thr Tyr Thr
 755 760 765
 Ser Ser Leu Gly Gly Lys Ile Pro Ile Arg Trp Thr Ala Pro Glu Ala
 770 775 780
 Ile Ala Tyr Arg Lys Phe Thr Ser Ala Ser Asp Val Trp Ser Tyr Gly
 785 790 795 800
 Ile Val Met Trp Glu Val Met Ser Tyr Gly Glu Arg Pro Tyr Trp Asp
 805 810 815
 Met Ser Asn Gln Asp Val Ile Asn Ala Val Glu Gln Asp Tyr Arg Leu
 820 825 830
 Pro Pro Pro Met Asp Cys Pro Thr Ala Leu His Gln Leu Met Leu Asp
 835 840 845
 Cys Trp Val Arg Asp Arg Asn Leu Arg Pro Lys Phe Ala Gln Ile Val
 850 855 860
 Asn Thr Leu Asp Lys Leu Ile Arg Asn Ala Ala Ser Leu Lys Val Ile
 865 870 875 880
 Ala Ser Val Gln Ser Gly Val Ser Gln Pro Leu Leu Asp Arg Thr Val
 885 890 895
 Pro Asp Tyr Thr Thr Phe Thr Thr Val Gly Asp Trp Leu Asp Ala Ile
 900 905 910
 Lys Met Gly Arg Tyr Lys Glu Asn Phe Val Asn Ala Gly Phe Ala Ser
 915 920 925
 Phe Asp Leu Val Ala Gln Met Thr Ala Glu Asp Leu Leu Arg Ile Gly
 930 935 940
 Val Thr Leu Ala Gly His Gln Lys Lys Ile Leu Ser Ser Ile Gln Asp
 945 950 955 960
 Met Arg Leu Gln Met Asn Gln Thr Leu Pro Val Gln Val
 965 970

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4097 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: linear

- (ix) FEATURE:
- (A) NAME/KEY: CDS
 - (B) LOCATION: 10..3042

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

71

CGGCTTCTG	ATG	CCC	GGC	CCG	GAG	CGC	ACC	ATG	GGG	CCG	TTG	TGG	TTC		48
Met	Pro	Gly	Pro	Glu	Arg	Thr	Met	Gly	Pro	Leu	Trp	Phe			
1				5					10						
TGC	TGT	TTG	CCC	CTC	GCC	CTC	TTG	CCT	CTG	CTC	GCC	GCC	GTG	GAA	GAG
Cys	Cys	Leu	Pro	Leu	Ala	Leu	Leu	Pro	Leu	Leu	Ala	Ala	Val	Glu	Glu
15				20							25				96
ACG	CTG	ATG	GAC	TCC	ACA	ACG	GCC	ACA	GCA	GAG	CTG	GGC	TGG	ATG	GTG
Thr	Leu	Met	Asp	Ser	Thr	Thr	Ala	Thr	Ala	Glu	Leu	Gly	Trp	Met	Val
30				35						40				45	144
CAT	CCT	CCC	TCA	GGG	TGG	GAA	GAG	GTG	AGT	GGA	TAC	GAT	GAG	AAC	ATG
His	Pro	Pro	Ser	Gly	Trp	Glu	Glu	Val	Ser	Gly	Tyr	Asp	Glu	Asn	Met
			50					55						60	192
AAC	ACC	ATC	CGC	ACC	TAC	CAG	GTG	TGC	AAC	GTC	TTT	GAA	TCC	AGC	CAA
Asn	Thr	Ile	Arg	Thr	Tyr	Gln	Val	Cys	Asn	Val	Phe	Glu	Ser	Ser	Gln
			65					70					75		240
AAC	AAC	TGG	CTG	CGG	ACC	AAG	TAC	ATC	CGG	AGG	CGA	GGA	GCG	CAC	CGC
Asn	Asn	Trp	Leu	Arg	Thr	Lys	Tyr	Ile	Arg	Arg	Arg	Gly	Ala	His	Arg
		80					85					90			288
ATC	CAC	GTG	GAG	ATG	AAA	TTC	TCC	GTT	CGG	GAC	TGC	AGC	AGC	ATC	CCC
Ile	His	Val	Glu	Met	Lys	Phe	Ser	Val	Arg	Asp	Cys	Ser	Ser	Ile	Pro
		95				100					105				336
AAC	GTC	CCG	GGC	TCC	TGT	AAG	GAG	ACT	TTT	AAC	CTC	TAT	TAC	TAC	GAA
Asn	Val	Pro	Gly	Ser	Cys	Lys	Glu	Thr	Phe	Asn	Leu	Tyr	Tyr	Tyr	Glu
110				115						120					384
TCA	GAC	TTT	GAC	TCT	GCC	ACC	AAG	ACT	TTT	CCT	AAC	TGG	ATG	GAA	AAC
Ser	Asp	Phe	Asp	Ser	Ala	Thr	Lys	Thr	Phe	Pro	Asn	Trp	Met	Glu	Asn
			130						135					140	432
CCT	TGG	ATG	AAG	GTA	GAT	ACA	ATT	GCT	GCC	GAC	GAG	AGC	TTC	TCG	CAG
Pro	Trp	Met	Lys	Val	Asp	Thr	Ile	Ala	Ala	Asp	Glu	Ser	Phe	Ser	Gln
			145					150					155		480
GTG	GAC	CTT	GGT	GGG	CGG	GTG	ATG	AAG	ATT	AAC	ACC	GAG	GTG	CGC	AGT
Val	Asp	Leu	Gly	Gly	Arg	Val	Met	Lys	Ile	Asn	Thr	Glu	Val	Arg	Ser
		160				165						170			528
TTT	GGG	CCT	GTC	TCC	AAA	AAC	GGT	TTC	TAC	CTG	GCC	TTC	CAG	GAC	TAC
Phe	Gly	Pro	Val	Ser	Lys	Asn	Gly	Phe	Tyr	Leu	Ala	Phe	Gln	Asp	Tyr
		175				180					185				576
GGG	GGC	TGC	ATG	TCC	TTG	ATT	GCA	GTC	CGT	GTC	TTT	TAC	CGC	AAG	TGT
Gly	Gly	Cys	Met	Ser	Leu	Ile	Ala	Val	Arg	Val	Phe	Tyr	Arg	Lys	Cys
190					195					200				205	624
CCC	CGT	GTG	ATC	CAG	AAC	GGG	GCG	GTC	TTC	CAG	GAA	ACC	CTC	TCG	GGA
Pro	Arg	Val	Ile	Gln	Asn	Gly	Ala	Val	Phe	Gln	Glu	Thr	Leu	Ser	Gly
			210						215					220	672
GCG	GAG	AGC	ACA	TCT	CTG	GTG	GCA	GCC	CGG	GGG	ACG	TGC	ATC	AGC	AAT
Ala	Glu	Ser	Thr	Ser	Leu	Val	Ala	Ala	Arg	Gly	Thr	Cys	Ile	Ser	Asn
			225					230					235		720
GCG	GAG	GAG	GTG	GAT	GTG	CCC	ATC	AAG	CTG	TAC	TGC	AAT	GGG	GAT	GGC
Ala	Glu	Glu	Val	Asp	Val	Pro	Ile	Lys	Leu	Tyr	Cys	Asn	Gly	Asp	Gly
		240				245						250			768
GAG	TGG	CTG	GTG	CCC	ATC	GGC	CGC	TGC	ATG	TGC	AGG	CCG	GGC	TAT	GAG
Glu	Trp	Leu	Val	Pro	Ile	Gly	Arg	Cys	Met	Cys	Arg	Pro	Gly	Tyr	Glu
		255				260					265				816

TCG Ser 270	GTG Val	GAG Glu	AAT Asn	GGG Gly	ACC Thr 275	GTC Val	TGC Cys	AGA Arg	GGC Gly	TGC Cys 280	CCA Pro	TCA Ser	GGG Gly	ACC Thr 285	TTC Phe	864
AAG Lys	GCC Ala	AGC Ser	CAA Gln	GGA Gly 290	GAT Asp	GAA Glu	GGA Gly	TGT Cys	GTC Val 295	CAT His	TGT Cys	CCA Pro	ATT Ile	AAC Asn 300	AGC Ser	912
CGG Arg	ACG Thr	ACT Thr	TCG Ser 305	GAA Glu	GGG Gly	GCC Ala	ACG Thr	AAC Asn 310	TGC Cys	GTG Val	TGC Cys	CGA Arg	AAC Asn 315	GGA Gly	TAT Tyr	960
TAC Tyr	CGG Arg	GCA Ala	GAT Asp	GCT Ala	GAC Asp	CCC Pro	GTC Val 325	GAC Asp	ATG Met	CCA Pro	TGC Cys	ACC Thr 330	ACC Thr	ATC Ile	CCA Pro	1008
TCT Ser	GCC Ala 335	CCC Pro	CAG Gln	GCC Ala	GTG Val 340	ATC Ile	TCC Ser	AGC Ser	GTG Val	AAT Asn 345	GAA Glu	ACC Thr	TCC Ser	CTG Leu	ATG Met	1056
CTG Leu 350	GAG Glu	TGG Trp	ACC Thr	CCA Pro 355	CCA Arg	CGA Asp	GAC Ser	TCA Ser	GGG Gly	GGC Arg 360	CGG Glu	GAG Arg	GAT Glu	CTG Asp	GTA Leu 365	1104
TAC Tyr	AAC Asn	ATC Ile	ATC Ile	TGC Cys 370	AAG Lys	AGC Ser	TGT Cys	GGG Gly	TCA Ser 375	GGC Gly	CGT Arg	GGG Gly	GCG Ala	TGC Cys 380	ACG Thr	1152
CGC Arg	TGT Cys	GGG Gly	GAC Asp 385	AAC Asn	GTG Val	CAG Gln	TTT Phe 390	GCC Ala	CCA Pro	CGC Arg	CAG Gln	CTG Leu 395	GGC Gly	CTG Leu	ACG Thr	1200
GAG Glu	CCT Pro 400	CGC Arg	ATC Ile	TAC Tyr	ATC Ile	AGC Ser	GAC Asp 405	CTG Leu	CTG Leu	GCC Ala	CAC His	ACG Thr 410	CAG Gln	TAC Tyr	ACC Thr	1248
TTT Phe 415	GAG Glu	ATC Ile	CAG Gln	GCT Ala	GTG Val	AAT Asn 420	GGG Gly	GTC Val	ACC Thr	GAC Asp 425	CAG Gln	AGC Ser	CCC Pro	TTC Phe	TCC Ser	1296
CCA Pro 430	CAG Gln	TTT Phe	GCA Ala	TCA Ser	GTG Val 435	AAT Asn	ATC Ile	ACC Thr	ACC Thr	AAC Asn 440	CAG Gln	GCT Ala	GCT Ala	CCT Pro	TCA Ser 445	1344
GCC Ala	GTG Val	TCC Ser	ATA Ile	ATG Met 450	CAC His	CAG Gln	GTC Val	AGC Ser	CGC Arg 455	ACT Thr	GTG Val	GAC Asp	AGC Ser	ATT Ile 460	ACC Thr	1392
CTC Leu	TCG Ser	TGG Trp	TCT Ser 465	CAA Gln	CCT Pro	GAC Asp	CAG Gln	CCC Pro	AAT Asn 470	GGA Gly	GTC Val	ATC Ile	CTG Leu 475	GAT Asp	TAT Tyr	1440
GAG Glu	CTG Leu	CAA Gln 480	TAC Tyr	TAT Tyr	GAG Glu	AAG Lys	AAC Asn 485	CTG Leu	AGT Ser	GAG Glu	TTA Leu	AAT Asn 490	TCA Ser	ACA Thr	GCA Ala	1488
GTG Val	AAG Lys 495	AGC Ser	CCC Pro	ACC Thr	AAC Asn	ACT Thr 500	GTG Val	ACA Thr	GTG Val	CAA Gln	AAC Asn 505	CTC Leu	AAA Lys	GCT Ala	GGC Gly	1536
ACC Thr 510	ATC Ile	TAT Tyr	GTC Val	TTC Phe 515	CAA Gln	GTG Val	CGA Arg	GCA Ala	CGT Arg	ACC Thr 520	GTG Val	GCT Ala	GGG Gly	TAT Tyr	GGC Gly 525	1584
CGG Arg	TAT Tyr	AGT Ser	GGC Gly	AAG Lys 530	ATG Met	TAC Tyr	TTC Phe	CAG Gln	ACC Thr 535	ATG Met	ACT Thr	GAA Glu	GCC Ala	GAG Glu	TAC Tyr 540	1632

CAG ACC AGT GTC CAG GAG AAG CTG CCA CTC ATC ATT GGC TCC TCT GCA Gln Thr Ser Val Gln Glu Lys Leu Pro Leu Ile Ile Gly Ser Ser Ala 545 550 555	1680
GCA GGA CTG GTG TTT CTC ATT GCT GTT GTC GTC ATC ATT ATT GTC TGC Ala Gly Leu Val Phe Leu Ile Ala Val Val Val Ile Ile Val Cys 560 565 570	1728
AAC AGA AGA CGG GGC TTT GAA CGT GCT GAC TCT GAG TAC ACT GAC AAG Asn Arg Arg Arg Gly Phe Glu Arg Ala Asp Ser Glu Tyr Thr Asp Lys 575 580 585	1776
CTG CAG CAC TAT ACC AGT GGC CAC AGT ACG TAC CGT GGT CCC CCG CCA Leu Gln His Tyr Thr Ser Gly His Ser Thr Tyr Arg Gly Pro Pro Pro 590 595 600 605	1824
GGC CTG GGG GTC CGC TCT CTC TTC GTG ACT CCA GGG ATG AAG ATT TAT Gly Leu Gly Val Arg Ser Leu Phe Val Thr Pro Gly Met Lys Ile Tyr 610 615 620	1872
ATC GAT CCA TTT ACC TAC GAA GAT CCC AAT GAG GCT GTC AGG GAA TTT Ile Asp Pro Phe Thr Tyr Glu Asp Pro Asn Glu Ala Val Arg Glu Phe 625 630 635	1920
GCA AAA GAA ATT GAT ATC TCC TGT GTG AAA ATC GAG CAG GTG ATT GGG Ala Lys Glu Ile Asp Ile Ser Cys Val Lys Ile Glu Gln Val Ile Gly 640 645 650	1968
GCA GGG GAG TTT GGT GAG GTG TGC AGT GGG CAT CTC AAG CTT CCT GGC Ala Gly Glu Phe Gly Glu Val Cys Ser Gly His Leu Lys Leu Pro Gly 655 660 665	2016
AAA AGA GAG ATC TTT GTG GCC ATC AAG ACC CTG AAG TCT GGT TAC ACA Lys Arg Glu Ile Phe Val Ala Ile Lys Thr Leu Lys Ser Gly Tyr Thr 670 675 680 685	2064
GAG AAG CAG AGA CGG GAC TTC CTG AGT GAA GCC AGC ATC ATG GGG CAG Glu Lys Gln Arg Arg Asp Phe Leu Ser Glu Ala Ser Ile Met Gly Gln 690 695 700	2112
TTT GAC CAC CCC AAT GTC ATC CAC CTG GAA GGG GTG GTG ACC AAG AGT Phe Asp His Pro Asn Val Ile His Leu Glu Gly Val Val Thr Lys Ser 705 710 715	2160
TCC CCA GTC ATG ATC ATT ACA GAG TTC ATG GAG AAT GGC TCG TTG GAC Ser Pro Val Met Ile Ile Thr Glu Phe Met Glu Asn Gly Ser Leu Asp 720 725 730	2208
TCC TTC TTG AGG CAA AAT GAT GGG CAG TTC ACA GTG ATC CAG CTG GTG Ser Phe Leu Arg Gln Asn Asp Gly Gln Phe Thr Val Ile Gln Leu Val 735 740 745	2256
GGC ATG TTG CGT GGC ATT GCA GCA GGC ATG AAG TAC CTG GCT GAT ATG Gly Met Leu Arg Gly Ile Ala Ala Gly Met Lys Tyr Leu Ala Asp Met 750 755 760 765	2304
AAC TAC GTG CAC CGG GAC CTG GCT GCC CGC AAC ATC CTG GTC AAC AGC Asn Tyr Val His Arg Asp Leu Ala Ala Arg Asn Ile Leu Val Asn Ser 770 775 780	2352
AAC CTG GTC TGC AAG GTG TCC GAC TTC GGC CTC TCC CGT TTC CTG GAG Asn Leu Val Cys Lys Val Ser Asp Phe Gly Leu Ser Arg Phe Leu Glu 785 790 795	2400
GAT GAC ACC TCT GAT CCC ACT TAC ACC AGC GCA CTG GGT GGA AAG ATC Asp Asp Thr Ser Asp Pro Thr Tyr Thr Ser Ala Leu Gly Gly Lys Ile 800 805 810	2448

CCA ATA CGG TGG ACA GCG CCT GAG GCA ATT CAG TAC CGA AAA TTC ACA Pro Ile Arg Trp Thr Ala Pro Glu Ala Ile Gln Tyr Arg Lys Phe Thr 815 820 825	2496
TCA GCC AGC GAT GTG TGG AGC TAT GGA ATA GTC ATG TGG GAG GTG ATG Ser Ala Ser Asp Val Trp Ser Tyr Gly Ile Val Met Trp Glu Val Met 830 835 840 845	2544
TCG TAC GGC GAG CGG CCT TAC TGG GAC ATG ACC AAT CAA GAT GTG ATA Ser Tyr Gly Glu Arg Pro Tyr Trp Asp Met Thr Asn Gln Asp Val Ile 850 855 860	2592
AAT GCT ATT GAG CAG GAC TAT CGG CTA CCA CCC CCT ATG GAT TGT CCA Asn Ala Ile Glu Gln Asp Tyr Arg Leu Pro Pro Pro Met Asp Cys Pro 865 870 875	2640
AAT GCC CTG CAC CAG CTA ATG CTT GAC TGC TGG CAG AAG GAT CGA AAC Asn Ala Leu His Gln Leu Met Leu Asp Cys Trp Gln Lys Asp Arg Asn 880 885 890	2688
CAC AGA CCC AAA TTT GGA CAG ATT GTC AAC ACT TTA GAC AAA ATG ATC His Arg Pro Lys Phe Gly Gln Ile Val Asn Thr Leu Asp Lys Met Ile 895 900 905	2736
CGA AAT CCT AAT AGT CTG AAA GCC ATG GCA CCT CTC TCC TCT GGG GTT Arg Asn Pro Asn Ser Leu Lys Ala Met Ala Pro Leu Ser Ser Gly Val 910 915 920 925	2784
AAC CTC CCT CTA CTT GAC CGC ACA ATC CCA GAT TAT ACC AGC TTC AAC Asn Leu Pro Leu Leu Asp Arg Thr Ile Pro Asp Tyr Thr Ser Phe Asn 930 935 940	2832
ACT GTG GAT GAA TGG CTG GAT GCC ATC AAG ATG AGC CAG TAC AAG GAG Thr Val Asp Glu Trp Leu Asp Ala Ile Lys Met Ser Gln Tyr Lys Glu 945 950 955	2880
AGC TTT GCC AGT GCT GGC TTC ACC ACC TTT GAT ATA GTA TCT CAG ATG Ser Phe Ala Ser Ala Gly Phe Thr Thr Phe Asp Ile Val Ser Gln Met 960 965 970	2928
ACT GTA GAG GAC ATT CTA CGA GTT GGG GTC ACT TTA GCA GGA CAC CAG Thr Val Glu Asp Ile Leu Arg Val Gly Val Thr Leu Ala Gly His Gln 975 980 985	2976
AAG AAA ATT CTG AAC AGT ATC CAG GTG ATG AGA GCA CAG ATG AAC CAA Lys Lys Ile Leu Asn Ser Ile Gln Val Met Arg Ala Gln Met Asn Gln 990 995 1000 1005	3024
ATT CAG TCT GTG GAG GTT TGATAGCAAC ACGTCCTCGT GCTCCACTTC Ile Gln Ser Val Glu Val 1010	3072
CTTGAGGCCC TGCTCCCTC TGCCCTGTG TGTCTGAGCT CCAGTTCTTG AGTGTCTGCTG	3132
GTGGATCAGA GACAGGCAGC TGCTCTGAGG ATCATGGCAA CAGGAAGAAA TGCCCTATCA	3192
TTGACAACGA GAAGTCATCA AGAGGTGAAA CAATGGAAAA CAATGGAAAA AGGGAACAAG	3252
TAAAGACAGC TATTTTGAAA ACCGAAAACA AACAGTGAAT TATTTTAAAA TAATAATAAA	3312
GCAATTGCAG TCTTGAAAAG GGCTCCAAGA CCAATGGGAG TCTCCAAAGG AAGAGAATAG	3372
AGCAGCTTCA TCTATTTCCT CTTACACAAG GGTGCTGCA GCTGGGCCCA GACACTTCTG	3432
GAGTAACGAG ACTTTTCAAG AAGATGAATG CAAAGAATGG TCACAAGAAG CACTTCTCTT	3492
TCTCACATGG GATGGCAGCT CTGGGAATGC CCGGCAGTCC TTCCTGAAAG CCCTGTTGGC	3552

75

AAATCGAAGA GGAGAGCCGA AGCTCTTTGG TGCTGTGGAA CCAAGTGCAT CTCAGAAATT 3612
 GTTGGACTTC TACAAAAGCT GAAGACATTC TTTTTTTTA AACAAGTAAA CTGATACTAG 3672
 AAGAGGCTGT TTCCTGCAAA TGAGAAGGAA TCTGTAACAC TGGCCCGGGG GGGGTGGGGA 3732
 ATGGGGGAAA TCAGTCCTTT TTACATCTCT TTATTTTCTC TTGTCATGGA ACAGTTTGTG 3792
 GAGTGACAGT TTCCTAAGGG TCCGTCCATC CACCCTCCAA TGGCATCATT GTTTCATACA 3852
 TATCATATGC ACAAGACTTA TAGTGATGTC CTCACTCGAT GCCAATGATC TTTCCCCAGA 3912
 AGACTTCCCA AGTACAGTAT GTAGTAGATT TTGATTACAA ATGCTGACGT GTACCTTTAT 3972
 TTTTCGGTTG TCGTTGTGG GAGATTCGTC CTTTACCTT GCTTTGTAA CACCAATTTG 4032
 TGAGTTTGGG GTTGGAATTT TTTGGTCGA TTGGGGTTGT TTTTTTTTTT TTTTTTTTTT 4092
 AACCG 4097

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1011 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Pro Gly Pro Glu Arg Thr Met Gly Pro Leu Trp Phe Cys Cys Leu
 1 5 10 15
 Pro Leu Ala Leu Leu Pro Leu Leu Ala Ala Val Glu Glu Thr Leu Met
 20 25 30
 Asp Ser Thr Thr Ala Thr Ala Glu Leu Gly Trp Met Val His Pro Pro
 35 40 45
 Ser Gly Trp Glu Glu Val Ser Gly Tyr Asp Glu Asn Met Asn Thr Ile
 50 55 60
 Arg Thr Tyr Gln Val Cys Asn Val Phe Glu Ser Ser Gln Asn Asn Trp
 65 70 75 80
 Leu Arg Thr Lys Tyr Ile Arg Arg Arg Gly Ala His Arg Ile His Val
 85 90 95
 Glu Met Lys Phe Ser Val Arg Asp Cys Ser Ser Ile Pro Asn Val Pro
 100 105 110
 Gly Ser Cys Lys Glu Thr Phe Asn Leu Tyr Tyr Tyr Glu Ser Asp Phe
 115 120 125
 Asp Ser Ala Thr Lys Thr Phe Pro Asn Trp Met Glu Asn Pro Trp Met
 130 135 140
 Lys Val Asp Thr Ile Ala Ala Asp Glu Ser Phe Ser Gln Val Asp Leu
 145 150 155 160
 Gly Gly Arg Val Met Lys Ile Asn Thr Glu Val Arg Ser Phe Gly Pro
 165 170 175
 Val Ser Lys Asn Gly Phe Tyr Leu Ala Phe Gln Asp Tyr Gly Gly Cys
 180 185 190

76

Met Ser Leu Ile Ala Val Arg Val Phe Tyr Arg Lys Cys Pro Arg Val
 195 200 205
 Ile Gln Asn Gly Ala Val Phe Gln Glu Thr Leu Ser Gly Ala Glu Ser
 210 215 220
 Thr Ser Leu Val Ala Ala Arg Gly Thr Cys Ile Ser Asn Ala Glu Glu
 225 230 235 240
 Val Asp Val Pro Ile Lys Leu Tyr Cys Asn Gly Asp Gly Glu Trp Leu
 245 250 255
 Val Pro Ile Gly Arg Cys Met Cys Arg Pro Gly Tyr Glu Ser Val Glu
 260 265 270
 Asn Gly Thr Val Cys Arg Gly Cys Pro Ser Gly Thr Phe Lys Ala Ser
 275 280 285
 Gln Gly Asp Glu Gly Cys Val His Cys Pro Ile Asn Ser Arg Thr Thr
 290 295 300
 Ser Glu Gly Ala Thr Asn Cys Val Cys Arg Asn Gly Tyr Tyr Arg Ala
 305 310 315 320
 Asp Ala Asp Pro Val Asp Met Pro Cys Thr Thr Ile Pro Ser Ala Pro
 325 330 335
 Gln Ala Val Ile Ser Ser Val Asn Glu Thr Ser Leu Met Leu Glu Trp
 340 345 350
 Thr Pro Pro Arg Asp Ser Gly Gly Arg Glu Asp Leu Val Tyr Asn Ile
 355 360 365
 Ile Cys Lys Ser Cys Gly Ser Gly Arg Gly Ala Cys Thr Arg Cys Gly
 370 375 380
 Asp Asn Val Gln Phe Ala Pro Arg Gln Leu Gly Leu Thr Glu Pro Arg
 385 390 395 400
 Ile Tyr Ile Ser Asp Leu Leu Ala His Thr Gln Tyr Thr Phe Glu Ile
 405 410 415
 Gln Ala Val Asn Gly Val Thr Asp Gln Ser Pro Phe Ser Pro Gln Phe
 420 425 430
 Ala Ser Val Asn Ile Thr Thr Asn Gln Ala Ala Pro Ser Ala Val Ser
 435 440 445
 Ile Met His Gln Val Ser Arg Thr Val Asp Ser Ile Thr Leu Ser Trp
 450 455 460
 Ser Gln Pro Asp Gln Pro Asn Gly Val Ile Leu Asp Tyr Glu Leu Gln
 465 470 475 480
 Tyr Tyr Glu Lys Asn Leu Ser Glu Leu Asn Ser Thr Ala Val Lys Ser
 485 490 495
 Pro Thr Asn Thr Val Thr Val Gln Asn Leu Lys Ala Gly Thr Ile Tyr
 500 505 510
 Val Phe Gln Val Arg Ala Arg Thr Val Ala Gly Tyr Gly Arg Tyr Ser
 515 520 525
 Gly Lys Met Tyr Phe Gln Thr Met Thr Glu Ala Glu Tyr Gln Thr Ser
 530 535 540

77

Val Gln Glu Lys Leu Pro Leu Ile Ile Gly Ser Ser Ala Ala Gly Leu
 545 550 555 560
 Val Phe Leu Ile Ala Val Val Val Ile Ile Ile Val Cys Asn Arg Arg
 565 570 575
 Arg Gly Phe Glu Arg Ala Asp Ser Glu Tyr Thr Asp Lys Leu Gln His
 580 585 590
 Tyr Thr Ser Gly His Ser Thr Tyr Arg Gly Pro Pro Pro Gly Leu Gly
 595 600 605
 Val Arg Ser Leu Phe Val Thr Pro Gly Met Lys Ile Tyr Ile Asp Pro
 610 615 620
 Phe Thr Tyr Glu Asp Pro Asn Glu Ala Val Arg Glu Phe Ala Lys Glu
 625 630 635 640
 Ile Asp Ile Ser Cys Val Lys Ile Glu Gln Val Ile Gly Ala Gly Glu
 645 650 655
 Phe Gly Glu Val Cys Ser Gly His Leu Lys Leu Pro Gly Lys Arg Glu
 660 665 670
 Ile Phe Val Ala Ile Lys Thr Leu Lys Ser Gly Tyr Thr Glu Lys Gln
 675 680 685
 Arg Arg Asp Phe Leu Ser Glu Ala Ser Ile Met Gly Gln Phe Asp His
 690 695 700
 Pro Asn Val Ile His Leu Glu Gly Val Val Thr Lys Ser Ser Pro Val
 705 710 715 720
 Met Ile Ile Thr Glu Phe Met Glu Asn Gly Ser Leu Asp Ser Phe Leu
 725 730 735
 Arg Gln Asn Asp Gly Gln Phe Thr Val Ile Gln Leu Val Gly Met Leu
 740 745 750
 Arg Gly Ile Ala Ala Gly Met Lys Tyr Leu Ala Asp Met Asn Tyr Val
 755 760 765
 His Arg Asp Leu Ala Ala Arg Asn Ile Leu Val Asn Ser Asn Leu Val
 770 775 780
 Cys Lys Val Ser Asp Phe Gly Leu Ser Arg Phe Leu Glu Asp Asp Thr
 785 790 795 800
 Ser Asp Pro Thr Tyr Thr Ser Ala Leu Gly Lys Ile Pro Ile Arg
 805 810 815
 Trp Thr Ala Pro Glu Ala Ile Gln Tyr Arg Lys Phe Thr Ser Ala Ser
 820 825 830
 Asp Val Trp Ser Tyr Gly Ile Val Met Trp Glu Val Met Ser Tyr Gly
 835 840 845
 Glu Arg Pro Tyr Trp Asp Met Thr Asn Gln Asp Val Ile Asn Ala Ile
 850 855 860
 Glu Gln Asp Tyr Arg Leu Pro Pro Pro Met Asp Cys Pro Asn Ala Leu
 865 870 875 880
 His Gln Leu Met Leu Asp Cys Trp Gln Lys Asp Arg Asn His Arg Pro
 885 890 895

Lys Phe Gly Gln Ile Val Asn Thr Leu Asp Lys Met Ile Arg Asn Pro
 900 905 910
 Asn Ser Leu Lys Ala Met Ala Pro Leu Ser Ser Gly Val Asn Leu Pro
 915 920 925
 Leu Leu Asp Arg Thr Ile Pro Asp Tyr Thr Ser Phe Asn Thr Val Asp
 930 935 940
 Glu Trp Leu Asp Ala Ile Lys Met Ser Gln Tyr Lys Glu Ser Phe Ala
 945 950 955 960
 Ser Ala Gly Phe Thr Thr Phe Asp Ile Val Ser Gln Met Thr Val Glu
 965 970 975
 Asp Ile Leu Arg Val Gly Val Thr Leu Ala Gly His Gln Lys Lys Ile
 980 985 990
 Leu Asn Ser Ile Gln Val Met Arg Ala Gln Met Asn Gln Ile Gln Ser
 995 1000 1005
 Val Glu Val
 1010

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3591 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..2965

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

C GGG GTC TCC TCG AGG GCG CGG CCG CCG GGC AGC AGC AGG AGC	46
Gly Val Ser Ser Arg Ala Arg Arg Pro Pro Gly Ser Ser Arg Ser	
1 5 10 15	
AGC AGG AGG GGG GTG ACC TCG GAG CTG GCA TGG ACA ACC CAT CCG GAG	94
Ser Arg Arg Gly Val Thr Ser Glu Leu Ala Trp Thr Thr His Pro Glu	
20 25 30	
ACG GGG TGG GAA GAG GTC AGT GGT TAC GAC GAG GCT ATG AAC CCC ATC	142
Thr Gly Trp Glu Glu Val Ser Gly Tyr Asp Glu Ala Met Asn Pro Ile	
35 40 45	
CGC ACA TAC CAG GTG TGC AAC GTG CGG GAG GCC AAC CAG AAC AAC TGG	190
Arg Thr Tyr Gln Val Cys Asn Val Arg Glu Ala Asn Gln Asn Asn Trp	
50 55 60	
CTT CGC ACC AAG TTC ATT CAG CGC CAG GAC GTC CAG CGT GTC TAC GTG	238
Leu Arg Thr Lys Phe Ile Gln Arg Gln Asp Val Gln Arg Val Tyr Val	
65 70 75	
GAG CTG AAA TTC ACT GTG CGG GAC TGC AAC AGC ATC CCC AAC ATC CCT	286
Glu Leu Lys Phe Thr Val Arg Asp Cys Asn Ser Ile Pro Asn Ile Pro	
80 85 90 95	
GGT TCC TGC AAA GAG ACC TTC AAC CTC TTC TAT TAT GAG TCA GAT ACG	334
Gly Ser Cys Lys Glu Thr Phe Asn Leu Phe Tyr Tyr Glu Ser Asp Thr	
100 105 110	

GAT TCT GCC TCT GCC AAT AGC CCT TTC TGG ATG GAG AAC CCC TAT ATC Asp Ser Ala Ser Ala Asn Ser Pro Phe Trp Met Glu Asn Pro Tyr Ile 115 120 125	382
AAA GTG GAT ACA ATT GCT CCG GAT GAG AGC TTC TCC AAA CTG GAG TCC Lys Val Asp Thr Ile Ala Pro Asp Glu Ser Phe Ser Lys Leu Glu Ser 130 135 140	430
GGC CGT GTG AAC ACC AAG GTG CGC AGC TTT GGG CCG CTC TCC AAG AAT Gly Arg Val Asn Thr Lys Val Arg Ser Phe Gly Pro Leu Ser Lys Asn 145 150 155	478
GGC TTT TAT CTG GCT TTC CAG GAC CTG GGG GCC TGC ATG TCC CTT ATC Gly Phe Tyr Leu Ala Phe Gln Asp Leu Gly Ala Cys Met Ser Leu Ile 160 165 170 175	526
TCC GTC CGG GCT TTC TAC AAG AAA TGT TCC AAC ACC ATC GCT GGC TTT Ser Val Arg Ala Phe Tyr Lys Lys Cys Ser Asn Thr Ile Ala Gly Phe 180 185 190	574
GCT ATC TTC CCG GAG ACC CTA ACG GGG GCT GAG CCC ACG TCG CTG GTC Ala Ile Phe Pro Glu Thr Leu Thr Gly Ala Glu Pro Thr Ser Leu Val 195 200 205	622
ATT GCG CCG GGC ACC TGC ATC CCC AAC GCA GTG GAA GTG TCT GTG CCC Ile Ala Pro Gly Thr Cys Ile Pro Asn Ala Val Glu Val Ser Val Pro 210 215 220	670
CTG AAG CTG TAC TGC AAC GGT GAT GGC GAG TGG ATG GTG CCT GTG GGA Leu Lys Leu Tyr Cys Asn Gly Asp Gly Glu Trp Met Val Pro Val Gly 225 230 235	718
GCG TGC ACG TGT GCT GCT GGG TAC GAG CCA GCC ATG AAG GAT ACC CAG Ala Cys Thr Cys Ala Ala Gly Tyr Glu Pro Ala Met Lys Asp Thr Gln 240 245 250 255	766
TGC CAA GCA TGC GGC CCG GGG ACG TTC AAA TCC AAG CAG GGC GAG GGC Cys Gln Ala Cys Gly Pro Gly Thr Phe Lys Ser Lys Gln Gly Glu Gly 260 265 270	814
CCC TGC TCC CCC TGC CCT CCC AAC AGC CGC ACC ACC GCG GGG GCA GCC Pro Cys Ser Pro Cys Pro Pro Asn Ser Arg Thr Thr Ala Gly Ala Ala 275 280 285	862
ACA GTC TGC ATA TGT CGC AGC GGC TTC TTC CGA GCA GAC GCG GAC CCC Thr Val Cys Ile Cys Arg Ser Gly Phe Phe Arg Ala Asp Ala Asp Pro 290 295 300	910
GCA GAC AGC GCC TGC ACC AGT GTG CCC TCA GCC CCA CGC AGC GTC ATC Ala Asp Ser Ala Cys Thr Ser Val Pro Ser Ala Pro Arg Ser Val Ile 305 310 315	958
TCC AAC GTG AAT GAG ACG TCG TTG GTG CTG GAG TGG AGC GAG CCG CAG Ser Asn Val Asn Glu Thr Ser Leu Val Leu Glu Trp Ser Glu Pro Gln 320 325 330 335	1006
GAC GCG GGC GGG CGG GAT GAC CTG CTC TAC AAC GTC ATC TGC AAG AAG Asp Ala Gly Gly Arg Asp Asp Leu Leu Tyr Asn Val Ile Cys Lys Lys 340 345 350	1054
TGC AGC GTG GAG CGG CGG CTG TGC AGC CGC TGC GAC GAC AAC GTG GAG Cys Ser Val Glu Arg Arg Leu Cys Ser Arg Cys Asp Asp Asn Val Glu 355 360 365	1102
TTC GTG CCG CGC CAG CTG GGC CTC ACT GGC CTC ACT GAG CGA CGC ATC Phe Val Pro Arg Gln Leu Gly Leu Thr Gly Leu Thr Glu Arg Arg Ile 370 375 380	1150

TAC Tyr 385	ATC Ile	AGC Ser	AAG Lys	GTG Val	ATG Met	GCC Ala 390	CAC His	CCC Pro	CAG Gln	TAC Tyr 395	ACC Thr	TTC Phe	GAG Glu	ATC Ile	CAG Gln	1198
GCG Ala 400	GTG Val	AAT Asn	GGC Gly	ATC Ile	TCC Ser 405	AGC Ser	AAG Lys	AGC Ser	CCC Pro	TAC Tyr 410	CCT Pro	CCC Pro	CAT His	TTT Phe	GCC Ala 415	1246
TCC Ser	GTC Val	AAC Asn	ATC Ile	ACG Thr 420	ACC Thr	AAC Asn	CAG Gln	GCA Ala	GCC Ala 425	CCA Pro	TCT Ser	GCC Ala	GTG Val	CCC Pro 430	ACC Thr	1294
ATG Met	CAT His	CTG Leu 435	CAC His	AGC Ser	AGC Ser	ACC Thr	GGG Gly 440	AAC Asn	AGC Ser	ATG Met	ACA Thr	CTG Leu	TCA Ser	TGG Trp 445	ACT Thr	1342
CCC Pro	CCG Pro	GAA Glu 450	AGG Arg	CCC Pro	AAC Asn	GGC Gly	ATC Ile 455	ATT Ile	CTC Leu	GAC Asp	TAT Tyr	GAA Glu 460	ATC Ile	AAG Lys	TAC Tyr	1390
TCC Ser 465	GAG Glu	AAG Lys	CAA Gln	GGC Gly	CAG Gln 470	GGT Gly	GAC Asp	GGC Gly	ATT Ile	GCC Ala	AAC Asn 475	ACT Thr	GTC Val	ACC Thr	AGC Ser	1438
CAG Gln 480	AAG Lys	AAC Asn	TCG Ser	GTG Val	CGG Arg 485	CTG Leu	GAC Asp	GGA Gly	CTG Leu	AAG Lys 490	GCC Ala	AAT Asn	GCT Ala	CGG Arg	TAC Tyr 495	1486
ATG Met	GTG Val	CAG Gln 500	GTC Val	CGG Arg	GCG Ala	CGC Arg	ACA Thr	GTG Val	GCT Ala 505	GGA Gly	TAC Tyr	GGC Gly	CGC Arg	TAC Tyr 510	AGC Ser	1534
CTC Leu	CCC Pro	ACC Thr	GAG Phe 515	TTC Gln	CAG Gln	ACG Thr	ACT Thr	GCG Ala 520	GAG Glu	GAT Asp	GGC Gly	TCC Ser	ACC Thr 525	AGC Ser	AAG Lys	1582
ACT Thr	TTC Phe	CAG Glu 530	GAG Leu	CTT Leu	CCT Pro	CTC Leu	ATC Ile 535	GTG Val	GGT Gly	TCA Ser	GCC Ala 540	ACC Thr	GCG Ala	GGA Gly	CTG Leu	1630
CTG Leu 545	TTT Phe	GTC Val	ATC Ile	GTG Val	GTG Val	GTC Val	ATC Ile	ATC Ile	GCT Ala	ATT Ile	GTC Val 555	TGC Cys	TTC Phe	AGG Arg	AAA Lys	1678
GGG Gly 560	ATG Met	GTT Val	ACT Thr	GAA Glu	CAA Gln 565	CTC Leu	CTC Leu	TCG Ser	TCT Ser	CCT Pro	TTG Leu 570	GGC Gly	AGG Arg	AAG Lys	CAG Gln 575	1726
CGC Arg	AAC Asn	AGC Ser	ACA Thr	GAT Asp 580	CCC Pro	GAG Glu	TAC Tyr	ACA Thr	GAG Lys	AAG Lys 585	CTG Leu	CAG Gln	CAA Gln	TAT Tyr 590	GTC Val	1774
ACT Thr	CCT Pro	GGG Gly 595	ATG Met	AAG Lys	GTC Val	TAC Tyr	ATT Ile	GAC Asp 600	CCC Pro	TTC Phe	ACC Thr	TAT Tyr	GAA Glu	GAC Asp 605	CCA Pro	1822
AAT Asn	GAA Glu	GCT Ala 610	GTC Val	CGG Arg	GAA Glu	TTC Phe	GCC Ala 615	AAA Lys	GAG Glu	ATT Ile	GAT Asp 620	ATC Ile	TCC Ser	TGT Cys	GTC Val	1870
AAA Lys 625	ATT Ile	GAG Glu	GAG Glu	GTC Val	ATT Ile	GGA Gly 630	GCA Ala	GGA Gly	GAG Glu	TTT Phe	GGT Gly 635	GAG Glu	GTG Val	TGC Cys	CGT Arg	1918
GGG Gly 640	CGC Arg	CTG Leu	AAG Lys	CTG Leu	CCT Pro	GGC Gly 645	CGC Arg	CGT Arg	GAG Glu	ATC Ile 650	TTT Phe	GTG Val	GCC Ala	ATC Ile	AAG Lys 655	1966

ACA CTG AAG GTG GGC TAC ACA GAG AGG CAG CGG CGG GAC TTC CTG AGT Thr Leu Lys Val Gly Tyr Thr Glu Arg Gln Arg Arg Asp Phe Leu Ser 660 665 670	2014
GAG GCC AGC ATC ATG GGC CAG TTC GAC CAC CCC AAC ATC ATC CAC CTG Glu Ala Ser Ile Met Gly Gln Phe Asp His Pro Asn Ile Ile His Leu 675 680 685	2062
GAG GGC GTG GTG ACC AAG AGC CGC CCT GTC ATG ATC ATC ACA GAG TTC Glu Gly Val Val Thr Lys Ser Arg Pro Val Met Ile Thr Glu Phe 690 695 700	2110
ATG GAG AAC TGC GCT CTC GAC TCC TTC CTC CGG CTG AAT GAT GGG CAG Met Glu Asn Cys Ala Leu Asp Ser Phe Leu Arg Leu Asn Asp Gly Gln 705 710 715	2158
TTC ACG GTC ATC CAG CTG GTG GGG ATG CTG CGA GGC ATC GCT GCT GGC Phe Thr Val Ile Gln Leu Val Gly Met Leu Arg Gly Ile Ala Ala Gly 720 725 730 735	2206
ATG AAG TAC CTC TCA GAG ATG AAC TAC GTG CAC CGA GAC CTG GCT GCC Met Lys Tyr Leu Ser Glu Met Asn Tyr Val His Arg Asp Leu Ala Ala 740 745 750	2254
CGC AAC ATC CTG GTC AAC AGC AAC TTG GTC TGC AAA GTG TCT GAC TTC Arg Asn Ile Leu Val Asn Ser Asn Leu Val Cys Lys Val Ser Asp Phe 755 760 765	2302
GGG CTC TCC CGC TTT TTG GAG GAT GAT CCA GCC GAC CCC ACC TAC ACC Gly Leu Ser Arg Phe Leu Glu Asp Asp Pro Ala Asp Pro Thr Tyr Thr 770 775 780	2350
AGC TCC CTG GGA GGC AAG ATC CCC ATC AGG TGG ACA GCT CCT GAG GCC Ser Ser Leu Gly Gly Lys Ile Pro Ile Arg Trp Thr Ala Pro Glu Ala 785 790 795	2398
ATC GCC TAC CGC AAA TTC ACG TCG GCC AGC GAC GTG TGG AGC TAC GGC Ile Ala Tyr Arg Lys Phe Thr Ser Ala Ser Asp Val Trp Ser Tyr Tyr 800 805 810 815	2446
ATC GTC ATG TGG GAA GTG ATG TCC TAC GGG GAG CGA CCC TAC TGG GAC Ile Val Met Trp Glu Val Met Ser Tyr Gly Glu Arg Pro Tyr Trp Asp 820 825 830	2494
ATG TCC AAC CAG GAT GTG ATC AAC GCG GTG GAG CAG GAT TAC CGC CTG Met Ser Asn Gln Asp Val Ile Asn Ala Val Glu Gln Asp Tyr Arg Leu 835 840 845	2542
CCA CCC CCC ATG GAC TGC CCC ACA GCA CTG CAC CAG CTG ATG CTG GAC Pro Pro Pro Met Asp Cys Pro Thr Ala Leu His Gln Leu Met Leu Asp 850 855 860	2590
TGC TGG GTG CGG GAC CGC AAC CTG CGG CCC AAG TTT GCA CAG ATT GTC Cys Trp Val Arg Asp Arg Asn Leu Arg Pro Lys Phe Ala Gln Ile Val 865 870 875	2638
AAC ACG CTG GAC AAG CTG ATC CGC AAT GCT GCC AGC CTG AAG GTC ATC Asn Thr Leu Asp Lys Leu Ile Arg Asn Ala Ala Ser Leu Lys Val Ile 880 885 890 895	2686
GCC AGC GTC CAG TCC GGT GTC TCC CAG CCG CTC CTG GAC CGC ACC GTG Ala Ser Val Gln Ser Gly Val Ser Gln Pro Leu Leu Asp Arg Thr Val 900 905 910	2734
CCC GAT TAC ACC ACC TTC ACC ACC GTG GGA GAC TGG CTG GAT GCC ATC Pro Asp Tyr Thr Thr Phe Thr Thr Val Gly Asp Trp Leu Asp Ala Ile 915 920 925	2782

82

AAA ATG GGA CGG TAC AAG GAG AAC TTC GTC AAC GCC GGC TTC GCC TCC Lys Met Gly Arg Tyr Lys Glu Asn Phe Val Asn Ala Gly Phe Ala Ser 930 935 940	2830
TTT GAC CTG GTG GCA CAG ATG ACA GCA GAG GAC CTG CTA AGG ATA GGA Phe Asp Leu Val Ala Gln Met Thr Ala Glu Asp Leu Leu Arg Ile Gly 945 950 955	2878
GTG ACG CTA GCA GGG CAC CAG AAG AAG ATC CTG AGC AGC ATT CAG GAC Val Thr Leu Ala Gly His Gln Lys Lys Ile Leu Ser Ser Ile Gln Asp 960 965 970 975	2926
ATG AGG CTG CAG ATG AAC CAG ACG CTG CCG GTT CAG GTT TGACCGCAGG Met Arg Leu Gln Met Asn Gln Thr Leu Pro Val Gln Val 980 985	2975
GACTCTGCAT TGGAACGGAC TGAGGGAACC TGCCAACCAG GTTCTGTTTG CGGTGCAGCC	3035
CGGCTTCCCG ATTTCCCTTT CCCGTGGCGC TCCTCTGCCT CGGACGCTCG CCGGGGACAG	3095
GCTGGGCCGG GCCACCTTC CCTGGATCAG AGGCACTCGT GCCGGGAGGG AGCCCGGCTT	3155
TTCGTCCCGT GTCCCGCAGC GCGGAGGCAG TGAACGCAGT CTTCATATTG AAGATGGATT	3215
ATGGGACGGA GATGGCGCAT CCGCTTCCCG CCCTGTCTCA GTGCTCATCA GTTTGAAGAG	3275
ATGTTCTGCT TCTTGGATTT CTTTACACCC CGGTTTTCCC CCCTCGAGTC CTCACCTCCC	3335
CCTATCCCTG AGGCCACAGA CTGTTGACCC GTCCGCTGAG TCCGTCAGAC GCTCCGAAGC	3395
CTTCCCGGAG CCCGGTCCCC GCGTGGAGAC GGC GCCAGG ACGGGGCTAC GGCCCCAGAC	3455
AATCACTCCA CCCCTCCGCA CGAGGGTCCT CACTGGGACG TGTCTGAAGG GGAAAGGCTC	3515
TGCTCCCTTT TTGGCTTTGC ACGCCAGAAC CCGAACCCCG TGAGATTTAC TATGCAGGGA	3575
GTTAGGCAAA AAAAAG	3591

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 988 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Gly Val Ser Ser Arg Ala Arg Arg Pro Pro Gly Ser Ser Arg Ser Ser 1 5 10 15
Arg Arg Gly Val Thr Ser Glu Leu Ala Trp Thr Thr His Pro Glu Thr 20 25 30
Gly Trp Glu Glu Val Ser Gly Tyr Asp Glu Ala Met Asn Pro Ile Arg 35 40 45
Thr Tyr Gln Val Cys Asn Val Arg Glu Ala Asn Gln Asn Asn Trp Leu 50 55 60
Arg Thr Lys Phe Ile Gln Arg Gln Asp Val Gln Arg Val Tyr Val Glu 65 70 75 80
Leu Lys Phe Thr Val Arg Asp Cys Asn Ser Ile Pro Asn Ile Pro Gly 85 90 95

83

Ser Cys Lys Glu Thr Phe Asn Leu Phe Tyr Tyr Glu Ser Asp Thr Asp
 100 105 110
 Ser Ala Ser Ala Asn Ser Pro Phe Trp Met Glu Asn Pro Tyr Ile Lys
 115 120 125
 Val Asp Thr Ile Ala Pro Asp Glu Ser Phe Ser Lys Leu Glu Ser Gly
 130 135 140
 Arg Val Asn Thr Lys Val Arg Ser Phe Gly Pro Leu Ser Lys Asn Gly
 145 150 155 160
 Phe Tyr Leu Ala Phe Gln Asp Leu Gly Ala Cys Met Ser Leu Ile Ser
 165 170 175
 Val Arg Ala Phe Tyr Lys Lys Cys Ser Asn Thr Ile Ala Gly Phe Ala
 180 185 190
 Ile Phe Pro Glu Thr Leu Thr Gly Ala Glu Pro Thr Ser Leu Val Ile
 195 200 205
 Ala Pro Gly Thr Cys Ile Pro Asn Ala Val Glu Val Ser Val Pro Leu
 210 215 220
 Lys Leu Tyr Cys Asn Gly Asp Gly Glu Trp Met Val Pro Val Gly Ala
 225 230 235 240
 Cys Thr Cys Ala Ala Gly Tyr Glu Pro Ala Met Lys Asp Thr Gln Cys
 245 250 255
 Gln Ala Cys Gly Pro Gly Thr Phe Lys Ser Lys Gln Gly Glu Gly Pro
 260 265 270
 Cys Ser Pro Cys Pro Pro Asn Ser Arg Thr Thr Ala Gly Ala Ala Thr
 275 280 285
 Val Cys Ile Cys Arg Ser Gly Phe Phe Arg Ala Asp Ala Asp Pro Ala
 290 295 300
 Asp Ser Ala Cys Thr Ser Val Pro Ser Ala Pro Arg Ser Val Ile Ser
 305 310 315 320
 Asn Val Asn Glu Thr Ser Leu Val Leu Glu Trp Ser Glu Pro Gln Asp
 325 330 335
 Ala Gly Gly Arg Asp Asp Leu Leu Tyr Asn Val Ile Cys Lys Lys Cys
 340 345 350
 Ser Val Glu Arg Arg Leu Cys Ser Arg Cys Asp Asp Asn Val Glu Phe
 355 360 365
 Val Pro Arg Gln Leu Gly Leu Thr Gly Leu Thr Glu Arg Arg Ile Tyr
 370 375 380
 Ile Ser Lys Val Met Ala His Pro Gln Tyr Thr Phe Glu Ile Gln Ala
 385 390 395 400
 Val Asn Gly Ile Ser Ser Lys Ser Pro Tyr Pro Pro His Phe Ala Ser
 405 410 415
 Val Asn Ile Thr Thr Asn Gln Ala Ala Pro Ser Ala Val Pro Thr Met
 420 425 430
 His Leu His Ser Ser Thr Gly Asn Ser Met Thr Leu Ser Trp Thr Pro
 435 440 445

84

Pro Glu Arg Pro Asn Gly Ile Ile Leu Asp Tyr Glu Ile Lys Tyr Ser
 450 455 460
 Glu Lys Gln Gly Gln Gly Asp Gly Ile Ala Asn Thr Val Thr Ser Gln
 465 470 475 480
 Lys Asn Ser Val Arg Leu Asp Gly Leu Lys Ala Asn Ala Arg Tyr Met
 485 490 495
 Val Gln Val Arg Ala Arg Thr Val Ala Gly Tyr Gly Arg Tyr Ser Leu
 500 505 510
 Pro Thr Glu Phe Gln Thr Thr Ala Glu Asp Gly Ser Thr Ser Lys Thr
 515 520 525
 Phe Gln Glu Leu Pro Leu Ile Val Gly Ser Ala Thr Ala Gly Leu Leu
 530 535 540
 Phe Val Ile Val Val Val Ile Ile Ala Ile Val Cys Phe Arg Lys Gly
 545 550 555 560
 Met Val Thr Glu Gln Leu Leu Ser Ser Pro Leu Gly Arg Lys Gln Arg
 565 570 575
 Asn Ser Thr Asp Pro Glu Tyr Thr Glu Lys Leu Gln Gln Tyr Val Thr
 580 585 590
 Pro Gly Met Lys Val Tyr Ile Asp Pro Phe Thr Tyr Glu Asp Pro Asn
 595 600 605
 Glu Ala Val Arg Glu Phe Ala Lys Glu Ile Asp Ile Ser Cys Val Lys
 610 615 620
 Ile Glu Glu Val Ile Gly Ala Gly Glu Phe Gly Glu Val Cys Arg Gly
 625 630 635 640
 Arg Leu Lys Leu Pro Gly Arg Arg Glu Ile Phe Val Ala Ile Lys Thr
 645 650 655
 Leu Lys Val Gly Tyr Thr Glu Arg Gln Arg Arg Asp Phe Leu Ser Glu
 660 665 670
 Ala Ser Ile Met Gly Gln Phe Asp His Pro Asn Ile Ile His Leu Glu
 675 680 685
 Gly Val Val Thr Lys Ser Arg Pro Val Met Ile Ile Thr Glu Phe Met
 690 695 700
 Glu Asn Cys Ala Leu Asp Ser Phe Leu Arg Leu Asn Asp Gly Gln Phe
 705 710 715 720
 Thr Val Ile Gln Leu Val Gly Met Leu Arg Gly Ile Ala Ala Gly Met
 725 730 735
 Lys Tyr Leu Ser Glu Met Asn Tyr Val His Arg Asp Leu Ala Ala Arg
 740 745 750
 Asn Ile Leu Val Asn Ser Asn Leu Val Cys Lys Val Ser Asp Phe Gly
 755 760 765
 Leu Ser Arg Phe Leu Glu Asp Asp Pro Ala Asp Pro Thr Tyr Thr Ser
 770 775 780
 Ser Leu Gly Gly Lys Ile Pro Ile Arg Trp Thr Ala Pro Glu Ala Ile
 785 790 795 800

85

Ala Tyr Arg Lys Phe Thr Ser Ala Ser Asp Val Trp Ser Tyr Gly Ile
805 810 815

Val Met Trp Glu Val Met Ser Tyr Gly Glu Arg Pro Tyr Trp Asp Met
820 825 830

Ser Asn Gln Asp Val Ile Asn Ala Val Glu Gln Asp Tyr Arg Leu Pro
835 840 845

Pro Pro Met Asp Cys Pro Thr Ala Leu His Gln Leu Met Leu Asp Cys
850 855 860

Trp Val Arg Asp Arg Asn Leu Arg Pro Lys Phe Ala Gln Ile Val Asn
865 870 875 880

Thr Leu Asp Lys Leu Ile Arg Asn Ala Ala Ser Leu Lys Val Ile Ala
885 890 895

Ser Val Gln Ser Gly Val Ser Gln Pro Leu Leu Asp Arg Thr Val Pro
900 905 910

Asp Tyr Thr Thr Phe Thr Thr Val Gly Asp Trp Leu Asp Ala Ile Lys
915 920 925

Met Gly Arg Tyr Lys Glu Asn Phe Val Asn Ala Gly Phe Ala Ser Phe
930 935 940

Asp Leu Val Ala Gln Met Thr Ala Glu Asp Leu Leu Arg Ile Gly Val
945 950 955 960

Thr Leu Ala Gly His Gln Lys Lys Ile Leu Ser Ser Ile Gln Asp Met
965 970 975

Arg Leu Gln Met Asn Gln Thr Leu Pro Val Gln Val
980 985

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3254 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 32..2980

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CGCTCTGCTC GCGCGTGCT GCCCGCCGA C ATG GAC CGC CGC CGC CTG CCG	52
Met Asp Arg Arg Arg Leu Pro	
1 5	
CTG CTG CTG CTC TGC GCT GCC CTC GGC TCC GCC GGG CGT CTG AGC GCC	100
Leu Leu Leu Leu Cys Ala Ala Leu Gly Ser Ala Gly Arg Leu Ser Ala	
10 15 20	
CGC CCC GGC AAC GAA GTT AAT CTG CTG GAT TCA AAA ACA ATT CAA GGG	148
Arg Pro Gly Asn Glu Val Asn Leu Leu Asp Ser Lys Thr Ile Gln Gly	
25 30 35	
GAG CTG GGC TGG ATC TCC TAC CCA TCA CAT GGG TGG GAA GAG ATT AGT	196
Glu Leu Gly Trp Ile Ser Tyr Pro Ser His Gly Trp Glu Glu Ile Ser	
40 45 50 55	

GGT GTT GAT GAG CAT TAT ACT CCA ATC AGA ACT TAC CAA GAG AGC AAT Gly Val Asp Glu His Tyr Thr Pro Ile Arg Thr Tyr Gln Glu Ser Asn 60 65 70	244
GTT ATG GAT CAC AGT CAA AAC AAT TGG CTG CGA ACA AAC TGG ATT CCA Val Met Asp His Ser Gln Asn Asn Trp Leu Arg Thr Asn Trp Ile Pro 75 80 85	292
CGC AAT TCA GCG CAG AAG ATA TAT GTG GAG CTC AAG TTT ACC TTG AGG Arg Asn Ser Ala Gln Lys Ile Tyr Val Glu Leu Lys Phe Thr Leu Arg 90 95 100	340
GAC TGC AAT AGT ATC CCT CTA GTT CTG GGC ACT TGC AAA GAG ACT TTC Asp Cys Asn Ser Ile Pro Leu Val Leu Gly Thr Cys Lys Glu Thr Phe 105 110 115	388
AAT CTG TAT TAC ATG GAA TCC GAT GAT GAC CAT TTG GCA AAG TTC AGA Asn Leu Tyr Tyr Met Glu Ser Asp Asp Asp His Leu Ala Lys Phe Arg 120 125 130 135	436
GAG CAC CAA TTT ACG AAG ATT GAC ACC ATG GCG GCT GAT GAG AGC TTC Glu His Gln Phe Thr Lys Ile Asp Thr Met Ala Ala Asp Glu Ser Phe 140 145 150	484
ACC CAG ATG GAT CTT GGG GAC CGG ATT CTC AAG CTG AAT ACC GAA GTC Thr Gln Met Asp Leu Gly Asp Arg Ile Leu Lys Leu Asn Thr Glu Val 155 160 165	532
CGC GAG GTG GGA CCT GTT AGT AAG AAG GGC TTT TAC TTG GCT TTC CAA Arg Glu Val Gly Pro Val Ser Lys Lys Gly Phe Tyr Leu Ala Phe Gln 170 175 180	580
GAT GTA GGT GCA TGT GTT GCC TTA GTC TCG GTG CGA GTG TAC TTC AAG Asp Val Gly Ala Cys Val Ala Leu Val Ser Val Arg Val Tyr Phe Lys 185 190 195	628
AAG TGC CCT TTC ACT GTC AAG AAC CTC GCC ATG TTT CCA GAT ACA GTT Lys Cys Pro Phe Thr Val Lys Asn Leu Ala Met Phe Pro Asp Thr Val 200 205 210 215	676
CCT ATG GAC TCC CAG TCC CTG GTG GAG GTG CGG GGT TCT TGT GTC AAT Pro Met Asp Ser Gln Ser Leu Val Glu Val Arg Gly Ser Cys Val Asn 220 225 230	724
CAT TCC AAG GAG GAA GAG CCA CCC AAG ATG TAC TGC AGC ACG GAA GGA His Ser Lys Glu Glu Glu Pro Pro Lys Met Tyr Cys Ser Thr Glu Gly 235 240 245	772
GAA TGG CTA GTG CCC ATA GGG AAG TGC TTG TGT AAT GCT GGC TAT GAA Glu Trp Leu Val Pro Ile Gly Lys Cys Leu Cys Asn Ala Gly Tyr Glu 250 255 260	820
GAG AGA GGC TTT GCG TGC CAA GCT TGT CGA CCT GGG TTC TAT AAA GCT Glu Arg Gly Phe Ala Cys Gln Ala Cys Arg Pro Gly Phe Tyr Lys Ala 265 270 275	868
TCT GCT GGC AAT GTG AAG TGT GCC AAA TGC CCA CCT CAC AGC TCT ACC Ser Ala Gly Asn Val Lys Cys Ala Lys Cys Pro Pro His Ser Ser Thr 280 285 290 295	916
TAT GAA GAT GCA TCT CTG AAC TGC AGG TGT GAA AAG AAT TAC TTT CGC Tyr Glu Asp Ala Ser Leu Asn Cys Arg Cys Glu Lys Asn Tyr Phe Arg 300 305 310	964
TCT GAG AAA GAC CCT CCA TCC ATG GCT TGC ACC AGA CCA CCA TCT GCT Ser Glu Lys Asp Pro Pro Ser Met Ala Cys Thr Arg Pro Pro Ser Ala 315 320 325	1012

CCA AGA AAC GTT ATT TCT AAC ATC AAT GAG ACA TCT GTT ATT CTG GAC Pro Arg Asn Val Ile Ser Asn Ile Asn Glu Thr Ser Val Ile Leu Asp 330 335 340	1060
TGG AGC TGG CCT CTT GAT ACA GGA GGT CGA AAA GAT GTC ACT TTC AAC Trp Ser Trp Pro Leu Asp Thr Gly Gly Arg Lys Asp Val Thr Phe Asn 345 350 355	1108
ATC ATT TGC AAA AAA TGT GGA GGA AGC AGC AAG ATA TGT GAG CCT TGC Ile Ile Cys Lys Lys Cys Gly Gly Ser Ser Lys Ile Cys Glu Pro Cys 360 365 370 375	1156
AGT GAC AAC GTA CGG TTC TTA CCC CGT CAG ACT GGC CTC ACC AAC ACC Ser Asp Asn Val Arg Phe Leu Pro Arg Gln Thr Gly Leu Thr Asn Thr 380 385 390	1204
ACG GTG ACA GTA GTG GAC CTT TTG GCA CAT ACC AAT TAC ACT TTT GAG Thr Val Thr Val Val Asp Leu Leu Ala His Thr Asn Tyr Thr Phe Glu 395 400 405	1252
ATT GAT GCA GTC AAC GGG GTA TCT GAC TTG AGT ACA CTT TCG AGA CAA Ile Asp Ala Val Asn Gly Val Ser Asp Leu Ser Thr Leu Ser Arg Gln 410 415 420	1300
TTT GCT GCT GTC AGC ATC ACG ACT AAT CAG GCT GCG CCA TCC CCC ATC Phe Ala Ala Val Ser Ile Thr Thr Asn Gln Ala Ala Pro Ser Pro Ile 425 430 435	1348
ACA GTG ATA AGG AAC GAC CGG ACA TCC AGG AAC AGC GTG TCT CTG TCT Thr Val Ile Arg Asn Asp Arg Thr Ser Arg Asn Ser Val Ser Leu Ser 440 445 450 455	1396
TGG CAG GAG CCT GAG CAC CCA AAT GGA ATC ATC TTG GAC TAC GAG GTC Trp Gln Glu Pro Glu His Pro Asn Gly Ile Ile Leu Asp Tyr Glu Val 460 465 470	1444
AAA TAC TAC GAA AAG CAG GAA CAA GAG ACA AGC TAT ACT ATT CTG AGA Lys Tyr Tyr Glu Lys Gln Glu Gln Thr Ser Tyr Thr Ile Leu Arg 475 480 485	1492
GCC AAA AGC ACT AAC GTT ACT ATC AGC GGC CTC AAA CCT GAT ACC ACC Ala Lys Ser Thr Asn Val Thr Ile Ser Gly Leu Lys Pro Asp Thr Thr 490 495 500	1540
TAC GTC TTC CAA ATT CGA GCC CGA ACT GCA GCT AGA TAT GGG ACA AGC Tyr Val Phe Gln Ile Arg Ala Arg Thr Ala Ala Arg Tyr Gly Thr Ser 505 510 515	1588
AGC CGC AAG TTT GAA TTT GAA ACC AGT CCA GAT TCA TTC TCC ATT TCC Ser Arg Lys Phe Glu Phe Glu Thr Ser Pro Asp Ser Phe Ser Ile Ser 520 525 530 535	1636
AGT GAA AAT AGC CAG GTC GTT ATG ATT GCC ATT TCA GCT GCA GTT GCC Ser Glu Asn Ser Gln Val Val Met Ile Ala Ile Ser Ala Ala Val Ala 540 545 550	1684
ATC ATT CTC CTC ACG GTT GTT GTG TAC GTC TTG ATT GGG AGA TTC TGC Ile Ile Leu Leu Thr Val Val Val Tyr Val Leu Ile Gly Arg Phe Cys 555 560 565	1732
GGA TAC AAG AAG TCT AAA CAT GGT ACC GAT GAG AAA AGA CTA CAT TTT Gly Tyr Lys Lys Ser Lys His Gly Thr Asp Glu Lys Arg Leu His Phe 570 575 580	1780
GGG AAT GGC CAC TTA AAA CTC CCA GGC CTG AGA ACT TAT GTA GAT CCA Gly Asn Gly His Leu Lys Leu Pro Gly Leu Arg Thr Tyr Val Asp Pro 585 590 595	1828

CAT His 600	ACG Thr	TAC Tyr	GAA Glu	GAT Asp	CCC Pro	AAT Asn	CAA Gln	GCT Ala	GTA Val	CAT His	GAA Glu	TTT Phe	GCC Ala	AAG Lys	GAA Glu	1876
CTA Leu	GAT Asp	GCT Ala	TCT Ser	AAT Asn	ATA Ile	TCA Ser	ATT Ile	GAT Asp	AAA Lys	GTT Val	GTT Val	GGA Gly	GCA Ala	GGG Gly	GAA Glu	1924
TTT Phe	GGA Gly	GAA Glu	GTG Val	TGC Cys	AGT Ser	GGG Gly	CGC Arg	CTG Leu	AAG Lys	CTG Leu	CCT Pro	TCT Ser	AAA Lys	AAG Lys	GAA Glu	1972
ATT Ile	TCA Ser	GTG Val	GCC Ala	ATC Ile	AAA Lys	ACT Thr	CTG Leu	AAA Lys	GCT Ala	GGC Gly	TAC Tyr	ACA Thr	GAA Glu	AAA Lys	CAG Gln	2020
AGA Arg	AGG Arg	GAT Asp	TTC Phe	CTG Leu	GGA Gly	GAA Glu	GCA Ala	AGC Ser	ATC Ile	ATG Met	GGG Gly	CAG Gln	TTT Phe	GAC Asp	CAC His	2068
CCC Pro	AAC Asn	ATC Ile	ATC Ile	CGA Arg	CTG Glu	GAG Gly	GGC Val	GTT Val	GTG Val	ACT Thr	AAA Lys	AGT Ser	AAA Lys	CCA Pro	GTT Val	2116
ATG Met	ATT Ile	GTT Val	ACT Thr	GAA Glu	TAC Tyr	ATG Met	GAA Glu	AAC Asn	GGT Gly	TCC Ser	TTG Leu	GAC Asp	AGC Ser	TTC Phe	CTA Leu	2164
CGG Arg	AAA Lys	CAT His	GAT Asp	GCC Ala	CAG Gln	TTC Phe	ACA Thr	GTC Val	ATT Ile	CAG Gln	CTA Leu	GTA Val	GGC Gly	ATG Met	CTT Leu	2212
CGT Arg	GGG Gly	ATC Ile	GCA Ala	TCT Ser	GGC Gly	ATG Met	AAA Lys	TAT Tyr	TTG Leu	TCA Ser	GAT Asp	ATG Met	GGT Gly	TAT Tyr	GTC Val	2260
CAC His	CGA Arg	GAT Asp	CTA Leu	GCT Ala	GCT Ala	CGT Arg	AAT Asn	ATA Ile	CTC Leu	ATC Ile	AAT Asn	AGT Ser	AAC Asn	TTG Leu	GTG Val	2308
TGC Cys	AAA Lys	GTC Val	TCA Ser	GAT Asp	TTT Phe	GGT Gly	CTT Leu	TCT Ser	CGT Arg	GTA Val	TTG Leu	GAA Glu	GAT Asp	GAC Asp	CCA Pro	2356
GAA Glu	GCT Ala	GCT Ala	TAC Tyr	ACA Thr	ACA Thr	AGG Arg	GGG Gly	GGC Gly	AAG Lys	ATT Ile	CCC Pro	ATC Ile	CGA Arg	TGG Trp	ACG Thr	2404
TCA Ser	CCA Pro	GAA Glu	GCC Ala	ATT Ile	GCA Ala	TAC Tyr	CGG Arg	AAG Lys	TTC Phe	ACA Thr	TCA Ser	GCC Ala	AGT Ser	GAT Asp	GCG Ala	2452
TGG Trp	AGC Ser	TAT Tyr	GGG Gly	ATT Ile	GTC Val	CTC Leu	TGG Trp	GAG Glu	GTG Val	ATG Met	TCT Ser	TAT Tyr	GGA Gly	GAA Glu	AGG Arg	2500
CCG Pro	TAC Tyr	TGG Trp	GAG Glu	ATG Met	TCC Ser	TTC Phe	CAG Gln	GAC Asp	GTA Val	ATT Ile	AAA Lys	GCC Ala	GTT Val	GAT Asp	GAA Glu	2548
GGG Gly	TAT Tyr	CGC Arg	TTG Leu	CCA Pro	CCT Pro	CCT Pro	ATG Met	GAC Asp	TGC Cys	CCA Pro	GCT Ala	GCC Ala	TTG Leu	TAT Tyr	CAG Gln	2596
CTG Leu	ATG Met	CTG Leu	GAC Asp	TGC Cys	TGG Trp	CAG Gln	AAA Lys	GAC Asp	AGA Arg	AAC Asn	AAC Asn	AGA Arg	CCC Pro	AAG Lys	TTT Phe	2644

89

GAG CAG ATT GTC AGC ATC CTG GAT AAG CTG ATC CGT AAT CCC AGC AGT Glu Gln Ile Val Ser Ile Leu Asp Lys Leu Ile Arg Asn Pro Ser Ser 875 880 885	2692
CTG AAA ATA ATC ACC AAT GCG GCA GCA AGG CCA TCA AAT CTT CTC CTG Leu Lys Ile Ile Thr Asn Ala Ala Arg Pro Ser Asn Leu Leu Leu 890 895 900	2740
GAC CAA AGT AAC ATT GAC ATT TCA GCG TTC CGC ACG GCA GGT GAT TGG Asp Gln Ser Asn Ile Asp Ile Ser Ala Phe Arg Thr Ala Gly Asp Trp 905 910 915	2788
CTC AAT GGT TTT CGA ACA GGA CAG TGC AAA GGC ATT TTC ACG GGT GTG Leu Asn Gly Phe Arg Thr Gly Gln Cys Lys Gly Ile Phe Thr Gly Val 920 925 930 935	2836
GAG TAC AGC TCC TGT GAT ACA ATA GCC AAG ATT TCC ACT GAT GAC ATG Glu Tyr Ser Ser Cys Asp Thr Ile Ala Lys Ile Ser Thr Asp Asp Met 940 945 950	2884
AAG AAA GTT GGT GTT ACA GTT GTG GGG CCT CAA AAG AAG ATT GTT AGC Lys Lys Val Gly Val Thr Val Val Gly Pro Gln Lys Lys Ile Val Ser 955 960 965	2932
AGT ATC AAA ACT CTA GAA ACT CAT ACG AAG AAC AGC CCT GTT CCT GTG Ser Ile Lys Thr Leu Glu Thr His Thr Lys Asn Ser Pro Val Pro Val 970 975 980	2980
TAAGGTACCA AAATGATGTT GCTGAGGACA GAAAAAAAAG AAAAGTCGCA TCAAAGTGCA	3040
AAAGCGATGG CTGATAACG GCACGGTTTA AAGGAGTTCT TTGCAGCAGT TTTGGAAACA	3100
TAATGGTTGA AATTTCAAAC CCACTGAGAC ACTCAAATAC TGAGTATAAA TGCCTTAAAA	3160
ATAGGAGCGA ACTTGTTTTT TATCTGTTAA TCCTGAAGGG TGGGTGCTCT TAACTGACTG	3220
TTAATGCAGA TAGTAAATTT CAAAAAAAAC AACG	3254

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 983 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met Asp Arg Arg Arg Leu Pro Leu Leu Leu Cys Ala Ala Leu Gly 1 5 10 15
Ser Ala Gly Arg Leu Ser Ala Arg Pro Gly Asn Glu Val Asn Leu Leu 20 25 30
Asp Ser Lys Thr Ile Gln Gly Glu Leu Gly Trp Ile Ser Tyr Pro Ser 35 40 45
His Gly Trp Glu Glu Ile Ser Gly Val Asp Glu His Tyr Thr Pro Ile 50 55 60
Arg Thr Tyr Gln Glu Ser Asn Val Met Asp His Ser Gln Asn Asn Trp 65 70 75 80
Leu Arg Thr Asn Trp Ile Pro Arg Asn Ser Ala Gln Lys Ile Tyr Val 85 90 95

Glu Leu Lys Phe Thr Leu Arg Asp Cys Asn Ser Ile Pro Leu Val Leu
 100 105 110
 Gly Thr Cys Lys Glu Thr Phe Asn Leu Tyr Tyr Met Glu Ser Asp Asp
 115 120 125
 Asp His Leu Ala Lys Phe Arg Glu His Gln Phe Thr Lys Ile Asp Thr
 130 135 140
 Met Ala Ala Asp Glu Ser Phe Thr Gln Met Asp Leu Gly Asp Arg Ile
 145 150 155 160
 Leu Lys Leu Asn Thr Glu Val Arg Glu Val Gly Pro Val Ser Lys Lys
 165 170 175
 Gly Phe Tyr Leu Ala Phe Gln Asp Val Gly Ala Cys Val Ala Leu Val
 180 185 190
 Ser Val Arg Val Tyr Phe Lys Lys Cys Pro Phe Thr Val Lys Asn Leu
 195 200 205
 Ala Met Phe Pro Asp Thr Val Pro Met Asp Ser Gln Ser Leu Val Glu
 210 215 220
 Val Arg Gly Ser Cys Val Asn His Ser Lys Glu Glu Glu Pro Pro Lys
 225 230 235 240
 Met Tyr Cys Ser Thr Glu Gly Glu Trp Leu Val Pro Ile Gly Lys Cys
 245 250 255
 Leu Cys Asn Ala Gly Tyr Glu Glu Arg Gly Phe Ala Cys Gln Ala Cys
 260 265 270
 Arg Pro Gly Phe Tyr Lys Ala Ser Ala Gly Asn Val Lys Cys Ala Lys
 275 280 285
 Cys Pro Pro His Ser Ser Thr Tyr Glu Asp Ala Ser Leu Asn Cys Arg
 290 295 300
 Cys Glu Lys Asn Tyr Phe Arg Ser Glu Lys Asp Pro Pro Ser Met Ala
 305 310 315 320
 Cys Thr Arg Pro Pro Ser Ala Pro Arg Asn Val Ile Ser Asn Ile Asn
 325 330 335
 Glu Thr Ser Val Ile Leu Asp Trp Ser Trp Pro Leu Asp Thr Gly Gly
 340 345 350
 Arg Lys Asp Val Thr Phe Asn Ile Ile Cys Lys Lys Cys Gly Gly Ser
 355 360 365
 Ser Lys Ile Cys Glu Pro Cys Ser Asp Asn Val Arg Phe Leu Pro Arg
 370 375 380
 Gln Thr Gly Leu Thr Asn Thr Thr Val Thr Val Val Asp Leu Leu Ala
 385 390 395 400
 His Thr Asn Tyr Thr Phe Glu Ile Asp Ala Val Asn Gly Val Ser Asp
 405 410 415
 Leu Ser Thr Leu Ser Arg Gln Phe Ala Ala Val Ser Ile Thr Thr Asn
 420 425 430
 Gln Ala Ala Pro Ser Pro Ile Thr Val Ile Arg Asn Asp Arg Thr Ser
 435 440 445

Arg Asn Ser Val Ser Leu Ser Trp Gln Glu Pro Glu His Pro Asn Gly
 450 455 460
 Ile Ile Leu Asp Tyr Glu Val Lys Tyr Tyr Glu Lys Gln Glu Gln Glu
 465 470 475 480
 Thr Ser Tyr Thr Ile Leu Arg Ala Lys Ser Thr Asn Val Thr Ile Ser
 485 490 495
 Gly Leu Lys Pro Asp Thr Thr Tyr Val Phe Gln Ile Arg Ala Arg Thr
 500 505 510
 Ala Ala Arg Tyr Gly Thr Ser Ser Arg Lys Phe Glu Phe Glu Thr Ser
 515 520 525
 Pro Asp Ser Phe Ser Ile Ser Ser Glu Asn Ser Gln Val Val Met Ile
 530 535 540
 Ala Ile Ser Ala Ala Val Ala Ile Ile Leu Leu Thr Val Val Val Tyr
 545 550 555 560
 Val Leu Ile Gly Arg Phe Cys Gly Tyr Lys Lys Ser Lys His Gly Thr
 565 570 575
 Asp Glu Lys Arg Leu His Phe Gly Asn Gly His Leu Lys Leu Pro Gly
 580 585 590
 Leu Arg Thr Tyr Val Asp Pro His Thr Tyr Glu Asp Pro Asn Gln Ala
 595 600 605
 Val His Glu Phe Ala Lys Glu Leu Asp Ala Ser Asn Ile Ser Ile Asp
 610 615 620
 Lys Val Val Gly Ala Gly Glu Phe Gly Glu Val Cys Ser Gly Arg Leu
 625 630 635 640
 Lys Leu Pro Ser Lys Lys Glu Ile Ser Val Ala Ile Lys Thr Leu Lys
 645 650 655
 Ala Gly Tyr Thr Glu Lys Gln Arg Arg Asp Phe Leu Gly Glu Ala Ser
 660 665 670
 Ile Met Gly Gln Phe Asp His Pro Asn Ile Ile Arg Leu Glu Gly Val
 675 680 685
 Val Thr Lys Ser Lys Pro Val Met Ile Val Thr Glu Tyr Met Glu Asn
 690 695 700
 Gly Ser Leu Asp Ser Phe Leu Arg Lys His Asp Ala Gln Phe Thr Val
 705 710 715 720
 Ile Gln Leu Val Gly Met Leu Arg Gly Ile Ala Ser Gly Met Lys Tyr
 725 730 735
 Leu Ser Asp Met Gly Tyr Val His Arg Asp Leu Ala Ala Arg Asn Ile
 740 745 750
 Leu Ile Asn Ser Asn Leu Val Cys Lys Val Ser Asp Phe Gly Leu Ser
 755 760 765
 Arg Val Leu Glu Asp Asp Pro Glu Ala Ala Tyr Thr Thr Arg Gly Gly
 770 775 780
 Lys Ile Pro Ile Arg Trp Thr Ser Pro Glu Ala Ile Ala Tyr Arg Lys
 785 790 795 800

Phe	Thr	Ser	Ala	Ser 805	Asp	Ala	Trp	Ser	Tyr	Gly	Ile	Val	Leu	Trp	Glu
Val	Met	Ser	Tyr 820	Gly	Glu	Arg	Pro	Tyr 825	Trp	Glu	Met	Ser	Phe 830	Gln	Asp
Val	Ile	Lys 835	Ala	Val	Asp	Glu	Gly 840	Tyr	Arg	Leu	Pro	Pro 845	Pro	Met	Asp
Cys	Pro	Ala	Ala	Leu	Tyr	Gln 855	Leu	Met	Leu	Asp	Cys 860	Trp	Gln	Lys	Asp
Arg 865	Asn	Asn	Arg	Pro	Lys 870	Phe	Glu	Gln	Ile	Val 875	Ser	Ile	Leu	Asp	Lys 880
Leu	Ile	Arg	Asn	Pro 885	Ser	Ser	Leu	Lys	Ile 890	Ile	Thr	Asn	Ala	Ala 895	Ala
Arg	Pro	Ser	Asn 900	Leu	Leu	Leu	Asp	Gln 905	Ser	Asn	Ile	Asp	Ile 910	Ser	Ala
Phe	Arg	Thr 915	Ala	Gly	Asp	Trp	Leu 920	Asn	Gly	Phe	Arg	Thr 925	Gly	Gln	Cys
Lys	Gly 930	Ile	Phe	Thr	Gly	Val 935	Glu	Tyr	Ser	Ser	Cys 940	Asp	Thr	Ile	Ala
Lys 945	Ile	Ser	Thr	Asp	Asp 950	Met	Lys	Lys	Val	Gly 955	Val	Thr	Val	Val	Gly 960
Pro	Gln	Lys	Lys	Ile 965	Val	Ser	Ser	Ile	Lys 970	Thr	Leu	Glu	Thr	His 975	Thr
Lys	Asn	Ser	Pro 980	Val	Pro	Val									

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4049 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: linear

(A) NAME/KEY: CDS
(B) LOCATION: 10..2994

CGGCTTCTG	ATG	CCC	GGC	CCG	GAG	CGC	ACC	ATG	GGG	CCG	TTG	TGG	TTC		48	
	Met	Pro	Gly	Pro	Glu	Arg	Thr	Met	Gly	Pro	Leu	Trp	Phe			
	1				5					10						
TGC	TGT	TTG	CCC	CTC	GCC	CTC	TTG	CCT	CTG	CTC	GCC	GCC	GTG	GAA	GAG	96
Cys	Cys	Leu	Pro	Leu	Ala	Leu	Leu	Pro	Leu	Leu	Ala	Ala	Val	Glu	Glu	
	15				20					25						
ACG	CTG	ATG	GAC	TCC	ACA	ACG	GCC	ACA	GCA	GAG	CTG	GGC	TGG	ATG	GTG	144
Thr	Leu	Met	Asp	Ser	Thr	Thr	Ala	Thr	Ala	Glu	Leu	Gly	Trp	Met	Val	
	30				35				40					45		
CAT	CCT	CCC	TCA	GGG	TGG	GAA	GAG	GTG	AGT	GGA	TAC	GAT	GAG	AAC	ATG	192
His	Pro	Pro	Ser	Gly	Trp	Glu	Glu	Val	Ser	Gly	Tyr	Asp	Glu	Asn	Met	
				50					55					60		

AAC ACC ATC CGC ACC TAC CAG GTG TGC AAC GTC TTT GAA TCC AGC CAA Asn Thr Ile Arg Thr Tyr Gln Val Cys Asn Val Phe Glu Ser Ser Gln 65 70 75	240
AAC AAC TGG CTG CGG ACC AAG TAC ATC CGG AGG CGA GGA GCG CAC CGC Asn Asn Trp Leu Arg Thr Lys Tyr Ile Arg Arg Arg Gly Ala His Arg 80 85 90	288
ATC CAC GTG GAG ATG AAA TTC TCC GTT CGG GAC TGC AGC AGC ATC CCC Ile His Val Glu Met Lys Phe Ser Val Arg Asp Cys Ser Ser Ile Pro 95 100 105	336
AAC GTC CCG GGC TCC TGT AAG GAG ACT TTT AAC CTC TAT TAC TAC GAA Asn Val Pro Gly Ser Cys Lys Glu Thr Phe Asn Leu Tyr Tyr Tyr Glu 110 115 120 125	384
TCA GAC TTT GAC TCT GCC ACC AAG ACT TTT CCT AAC TGG ATG GAA AAC Ser Asp Phe Asp Ser Ala Thr Lys Thr Phe Pro Asn Trp Met Glu Asn 130 135 140	432
CCT TGG ATG AAG GTA GAT ACA ATT GCT GCC GAC GAG AGC TTC TCG CAG Pro Trp Met Lys Val Asp Thr Ile Ala Ala Asp Glu Ser Phe Ser Gln 145 150 155	480
GTG GAC CTT GGT GGG CGG GTG ATG AAG ATT AAC ACC GAG GTG CGC AGT Val Asp Leu Gly Gly Arg Val Met Lys Ile Asn Thr Glu Val Arg Ser 160 165 170	528
TTT GGG CCT GTC TCC AAA AAC GGT TTC TAC CTG GCC TTC CAG GAC TAC Phe Gly Pro Val Ser Lys Asn Gly Phe Tyr Leu Ala Phe Gln Asp Tyr 175 180 185	576
GGG GGC TGC ATG TCC TTG ATT GCA GTC CGT GTC TTT TAC CGC AAG TGT Gly Gly Cys Met Ser Lys Ile Ala Val Arg Val Phe Tyr Arg Lys Cys 190 195 200 205	624
CCC CGT GTG ATC CAG AAC GGG GCG GTC TTC CAG GAA ACC CTC TCG GGA Pro Arg Val Ile Gln Asn Gly Ala Val Phe Gln Glu Thr Leu Ser Gly 210 215 220	672
GCG GAG AGC ACA TCT CTG GTG GCA GCC CGG GGG ACG TGC ATC AGC AAT Ala Glu Ser Thr Ser Leu Val Ala Ala Arg Gly Thr Cys Ile Ser Asn 225 230 235	720
GCG GAG GAG GTG GAT GTG CCC ATC AAG CTG TAC TGC AAT GGG GAT GGC Ala Glu Glu Val Asp Val Pro Ile Lys Leu Tyr Cys Asn Gly Asp Gly 240 245 250	768
GAG TGG CTG GTG CCC ATC GGC CGC TGC ATG TGC AGG CCG GGC TAT GAG Glu Trp Leu Val Pro Ile Gly Arg Cys Met Cys Arg Pro Gly Tyr Glu 255 260 265	816
TCG GTG GAG AAT GGG ACC GTC TGC AGA GGC TGC CCA TCA GGG ACC TTC Ser Val Glu Asn Gly Thr Val Cys Arg Gly Cys Pro Ser Gly Thr Phe 270 275 280 285	864
AAG GCC AGC CAA GGA GAT GAA GGA TGT GTC CAT TGT CCA ATT AAC AGC Lys Ala Ser Gln Gly Asp Glu Gly Cys Val His Cys Pro Ile Asn Ser 290 295 300	912
CGG ACG ACT TCG GAA GGG GCC ACG AAC TGC GTG TGC CGA AAC GGA TAT Arg Thr Thr Ser Glu Gly Ala Thr Asn Cys Val Cys Arg Asn Gly Tyr 305 310 315	960
TAC CGG GCA GAT GCT GAC CCC GTC GAC ATG CCA TGC ACC ACC ATC CCA Tyr Arg Ala Asp Ala Asp Pro Val Asp Met Pro Cys Thr Thr Ile Pro 320 325 330	1008

TCT	GCC	CCC	CAG	GCC	GTG	ATC	TCC	AGC	GTG	AAT	GAA	ACC	TCC	CTG	ATG	1056
Ser	Ala	Pro	Gln	Ala	Val	Ile	Ser	Ser	Val	Asn	Glu	Thr	Ser	Leu	Met	
335						340					345					
CTG	GAG	TGG	ACC	CCA	CCA	CGA	GAC	TCA	GGG	GGC	CGG	GAG	GAT	CTG	GTA	1104
Leu	Glu	Trp	Thr	Pro	Pro	Arg	Asp	Ser	Gly	Gly	Arg	Glu	Asp	Leu	Val	
350					355					360					365	
TAC	AAC	ATC	ATC	TGC	AAG	AGC	TGT	GGG	TCA	GGC	CGT	GGG	GCG	TGC	ACG	1152
Tyr	Asn	Ile	Ile	Cys	Lys	Ser	Cys	Gly	Ser	Gly	Arg	Gly	Ala	Cys	Thr	
				370					375					380		
CGC	TGT	GGG	GAC	AAC	GTG	CAG	TTT	GCC	CCA	CGC	CAG	CTG	GGC	CTG	ACG	1200
Arg	Cys	Gly	Asp	Asn	Val	Gln	Phe	Ala	Pro	Arg	Gln	Leu	Gly	Leu	Thr	
			385					390					395			
GAG	CCT	CGC	ATC	TAC	ATC	AGC	GAC	CTG	CTG	GCC	CAC	ACG	CAG	TAC	ACC	1248
Glu	Pro	Arg	Ile	Tyr	Ile	Ser	Asp	Leu	Leu	Ala	His	Thr	Gln	Tyr	Thr	
		400					405					410				
TTT	GAG	ATC	CAG	GCT	GTG	AAT	GGG	GTC	ACC	GAC	CAG	AGC	CCC	TTC	TCC	1296
Phe	Glu	Ile	Gln	Ala	Val	Asn	Gly	Val	Thr	Asp	Gln	Ser	Pro	Phe	Ser	
415						420					425					
CCA	CAG	TTT	GCA	TCA	GTG	AAT	ATC	ACC	ACC	AAC	CAG	GCT	GCT	CCT	TCA	1344
Pro	Gln	Phe	Ala	Ser	Val	Asn	Ile	Thr	Thr	Asn	Gln	Ala	Ala	Pro	Ser	
430					435					440					445	
GCC	GTG	TCC	ATA	ATG	CAC	CAG	GTC	AGC	CGC	ACT	GTG	GAC	AGC	ATT	ACC	1392
Ala	Val	Ser	Ile	Met	His	Gln	Val	Ser	Arg	Thr	Val	Asp	Ser	Ile	Thr	
				450					455					460		
CTC	TCG	TGG	TCT	CAA	CCT	GAC	CAG	CCC	AAT	GGA	GTC	ATC	CTG	GAT	TAT	1440
Leu	Ser	Trp	Ser	Gln	Pro	Asp	Gln	Pro	Asn	Gly	Val	Ile	Leu	Asp	Tyr	
			465					470					475			
GAG	CTG	CAA	TAC	TAT	GAG	AAG	AAC	CTG	AGT	GAG	TTA	AAT	TCA	ACA	GCA	1488
Glu	Leu	Gln	Tyr	Tyr	Glu	Lys	Asn	Leu	Ser	Glu	Leu	Asn	Ser	Thr	Ala	
		480					485					490				
GTG	AAG	AGC	CCC	ACC	AAC	ACT	GTG	ACA	GTG	CAA	AAC	CTC	AAA	GCT	GGC	1536
Val	Lys	Ser	Pro	Thr	Asn	Thr	Val	Thr	Val	Gln	Asn	Leu	Lys	Ala	Gly	
	495					500					505					
ACC	ATC	TAT	GTC	TTC	CAA	GTG	CGA	GCA	CGT	ACC	GTG	GCT	GGG	TAT	GGC	1584
Thr	Ile	Tyr	Val	Phe	Gln	Val	Arg	Ala	Arg	Thr	Val	Ala	Gly	Tyr	Gly	
510					515					520					525	
CGG	TAT	AGT	GGC	AAG	ATG	TAC	TTC	CAG	ACC	ATG	ACT	GAA	GCC	GAG	TAC	1632
Arg	Tyr	Ser	Gly	Lys	Met	Tyr	Phe	Gln	Thr	Met	Thr	Glu	Ala	Glu	Tyr	
				530					535					540		
CAG	ACC	AGT	GTC	CAG	GAG	AAG	CTG	CCA	CTC	ATC	ATT	GGC	TCC	TCT	GCA	1680
Gln	Thr	Ser	Val	Gln	Glu	Lys	Leu	Pro	Leu	Ile	Ile	Gly	Ser	Ser	Ala	
			545					550					555			
GCA	GGA	CTG	GTG	TTT	CTC	ATT	GCT	GTT	GTC	GTC	ATC	ATT	ATT	GTC	TGC	1728
Ala	Gly	Leu	Val	Phe	Leu	Ile	Ala	Val	Val	Val	Ile	Ile	Ile	Val	Cys	
		560					565					570				
AAC	AGA	AGA	CGG	GGC	TTT	GAA	CGT	GCT	GAC	TCT	GAG	TAC	ACT	GAC	AAG	1776
Asn	Arg	Arg	Arg	Gly	Phe	Glu	Arg	Ala	Asp	Ser	Glu	Tyr	Thr	Asp	Lys	
	575					580					585					
CTG	CAG	CAC	TAT	ACC	AGT	GGC	CAC	ATG	ACT	CCA	GGG	ATG	AAG	ATT	TAT	1824
Leu	Gln	His	Tyr	Thr	Ser	Gly	His	Met	Thr	Pro	Gly	Met	Lys	Ile	Tyr	
590						595				600					605	

ATC GAT CCA TTT ACC TAC GAA GAT CCC AAT GAG GCT GTC AGG GAA TTT Ile Asp Pro Phe Thr Tyr Glu Asp Pro Asn Glu Ala Val Arg Glu Phe 610 615 620	1872
GCA AAA GAA ATT GAT ATC TCC TGT GTG AAA ATC GAG CAG GTG ATT GGG Ala Lys Glu Ile Asp Ile Ser Cys Val Lys Ile Glu Gln Val Ile Gly 625 630 635	1920
GCA GGG GAG TTT GGT GAG GTG TGC AGT GGG CAT CTC AAG CTT CCT GGC Ala Gly Glu Phe Gly Glu Val Cys Ser Gly His Leu Lys Leu Pro Gly 640 645 650	1968
AAA AGA GAG ATC TTT GTG GCC ATC AAG ACC CTG AAG TCT GGT TAC ACA Lys Arg Glu Ile Phe Val Ala Ile Lys Thr Leu Lys Ser Gly Tyr Thr 655 660 665	2016
GAG AAG CAG AGA CGG GAC TTC CTG AGT GAA GCC AGC ATC ATG GGG CAG Glu Lys Gln Arg Arg Asp Phe Leu Ser Glu Ala Ser Ile Met Gly Gln 670 675 680 685	2064
TTT GAC CAC CCC AAT GTC ATC CAC CTG GAA GGG GTG GTG ACC AAG AGT Phe Asp His Pro Asn Val Ile His Leu Glu Gly Val Val Thr Lys Ser 690 695 700	2112
TCC CCA GTC ATG ATC ATT ACA GAG TTC ATG GAG AAT GGC TCG TTG GAC Ser Pro Val Met Ile Ile Thr Glu Phe Met Glu Asn Gly Ser Leu Asp 705 710 715	2160
TCC TTC TTG AGG CAA AAT GAT GGG CAG TTC ACA GTG ATC CAG CTG GTG Ser Phe Leu Arg Gln Asn Asp Gly Gln Phe Thr Val Ile Gln Leu Val 720 725 730	2208
GGC ATG TTG CGT GGC ATT GCA GCA GGC ATG AAG TAC CTG GCT GAT ATG Gly Met Leu Arg Gly Ile Ala Ala Gly Met Lys Tyr Leu Ala Asp Met 735 740 745	2256
AAC TAC GTG CAC CGG GAC CTG GCT GCC CGC AAC ATC CTG GTC AAC AGC Asn Tyr Val His Arg Asp Leu Ala Ala Arg Asn Ile Leu Val Asn Ser 750 755 760 765	2304
AAC CTG GTC TGC AAG GTG TCC GAC TTC GGC CTC TCC CGT TTC CTG GAG Asn Leu Val Cys Lys Val Ser Asp Phe Gly Leu Ser Arg Phe Leu Glu 770 775 780	2352
GAT GAC ACC TCT GAT CCC ACT TAC ACC AGC GCA CTG GGT GGA AAG ATC Asp Asp Thr Ser Asp Pro Thr Tyr Thr Ser Ala Leu Gly Gly Lys Ile 785 790 795	2400
CCA ATA CGG TGG ACA GCG CCT GAG GCA ATT CAG TAC CGA AAA TTC ACA Pro Ile Arg Trp Thr Ala Pro Glu Ala Ile Gln Tyr Arg Lys Phe Thr 800 805 810	2448
TCA GCC AGC GAT GTG TGG AGC TAT GGA ATA GTC ATG TGG GAG GTG ATG Ser Ala Ser Asp Val Trp Ser Tyr Gly Ile Val Met Trp Glu Val Met 815 820 825	2496
TCG TAC GGC GAG CGG CCT TAC TGG GAC ATG ACC AAT CAA GAT GTG ATA Ser Tyr Gly Glu Arg Pro Tyr Trp Asp Met Thr Asn Gln Asp Val Ile 830 835 840 845	2544
AAT GCT ATT GAG CAG GAC TAT CGG CTA CCA CCC CCT ATG GAT TGT CCA Asn Ala Ile Glu Gln Asp Tyr Arg Leu Pro Pro Pro Met Asp Cys Pro 850 855 860	2592
AAT GCC CTG CAC CAG CTA ATG CTT GAC TGC TGG CAG AAG GAT CGA AAC Asn Ala Leu His Gln Leu Met Leu Asp Cys Trp Gln Lys Asp Arg Asn 865 870 875	2640

CAC AGA CCC AAA TTT GGA CAG ATT GTC AAC ACT TTA GAC AAA ATG ATC His Arg Pro Lys Phe Gly Gln Ile Val Asn Thr Leu Asp Lys Met Ile 880 885 890	2688
CGA AAT CCT AAT AGT CTG AAA GCC ATG GCA CCT CTC TCC TCT GGG GTT Arg Asn Pro Asn Ser Leu Lys Ala Met Ala Pro Leu Ser Ser Gly Val 895 900 905	2736
AAC CTC CCT CTA CTT GAC CGC ACA ATC CCA GAT TAT ACC AGC TTC AAC Asn Leu Pro Leu Leu Asp Arg Thr Ile Pro Asp Tyr Thr Ser Phe Asn 910 915 920 925	2784
ACT GTG GAT GAA TGG CTG GAT GCC ATC AAG ATG AGC CAG TAC AAG GAG Thr Val Asp Glu Trp Leu Asp Ala Ile Lys Met Ser Gln Tyr Lys Glu 930 935 940	2832
AGC TTT GCC AGT GCT GGC TTC ACC ACC TTT GAT ATA GTA TCT CAG ATG Ser Phe Ala Ser Ala Gly Phe Thr Thr Phe Asp Ile Val Ser Gln Met 945 950 955	2880
ACT GTA GAG GAC ATT CTA CGA GTT GGG GTC ACT TTA GCA GGA CAC CAG Thr Val Glu Asp Ile Leu Arg Val Gly Val Thr Leu Ala Gly His Gln 960 965 970	2928
AAG AAA ATT CTG AAC AGT ATC CAG GTG ATG AGA GCA CAG ATG AAC CAA Lys Lys Ile Leu Asn Ser Ile Gln Val Met Arg Ala Gln Met Asn Gln 975 980 985	2976
ATT CAG TCT GTG GAG GTT TGATAGCAAC ACGTCCTCGT GCTCCACTTC Ile Gln Ser Val Glu Val 990 995	3024
CTTGAGGCCC TGCTCCCCTC TGCCCTGTG TGTCTGAGCT CCAGTTCTTG AGTGTCTGTC	3084
GTGGATCAGA GACAGGCAGC TGCTCTGAGG ATCATGGCAA CAGGAAGAAA TGCCCTATCA	3144
TTGACAACGA GAAGTCATCA AGAGGTGAAA CAATGGAAAA CAATGGAAAA AGGGAACAAG	3204
TAAAGACAGC TATTTTGAAA ACCGAAAACA AACAGTGAAT TATTTTTAAA TAATAATAAA	3264
GCAATTGCAG TCTTGAAAAG GGCTCCAAGA CCAATGGGAG TCTCCAAGG AAGAGAATAG	3324
AGCAGCTTCA TCTATTTCTT CTTACACAAG GGTGTGCTCA GCTGGGCCCC GACACTTCTG	3384
GAGTAACGAG ACTTTTCAAG AAGATGAATG CAAAGAATGG TCACAAGAAG CACTTCTCTT	3444
TCTCATATGG GATGGCAGCT CTGGAATGC CCGGCAGTCC TTCCTGAAAG CCCTGTGGC	3504
AAATCGAAGA GGAGAGCCGA AGCTCTTTGG TGCTGTGGAA CCAAGTGCAT CTCAGAAATT	3564
GTTGGACTTC TACAAAAGCT GAAGACATTC TTTTTTTTTA AACAAGTAAA CTGATACTAG	3624
AAGAGGCTGT TTCCGTCAAA TGAGAAGGAA TCTGTAACAC TGGCCCGGGG GGGGTGGGGA	3684
ATGGGGGAAA TCAGTCCTTT TTACATCTCT TTATTTTCTC TTGTCATGGA ACAGTTTGT	3744
GAGTGACAGT TTCCTAAGGG TCCGTCCATC CACCCCTCAA TGGCATCATT GTTTCATACA	3804
TATCATATGC ACAAGACTTA TAGTGATGTC CTCACTCGAT GCCAATGATC TTTCCCCAGA	3864
AGACTTCCCA AGTACAGTAT GTAGTAGATT TTGATTACAA ATGCTGACGT GTACCTTTAT	3924
TTTTCGGTTG TCGTTGTTGG GAGATTCGTC CTTTACCTT GCTTTGTAA CACCAATTG	3984
TGAGTTTGGG GTTGAATTT TTTGGTCTGA TTGGGGTTGT TTTTTTTTTT TTTTTTTTTT	4044
AACCG	4049

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 995 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```

Met Pro Gly Pro Glu Arg Thr Met Gly Pro Leu Trp Phe Cys Cys Leu
 1             5             10             15
Pro Leu Ala Leu Leu Pro Leu Leu Ala Ala Val Glu Glu Thr Leu Met
          20             25             30
Asp Ser Thr Thr Ala Thr Ala Glu Leu Gly Trp Met Val His Pro Pro
          35             40             45
Ser Gly Trp Glu Glu Val Ser Gly Tyr Asp Glu Asn Met Asn Thr Ile
          50             55             60
Arg Thr Tyr Gln Val Cys Asn Val Phe Glu Ser Ser Gln Asn Asn Trp
          65             70             75             80
Leu Arg Thr Lys Tyr Ile Arg Arg Arg Gly Ala His Arg Ile His Val
          85             90             95
Glu Met Lys Phe Ser Val Arg Asp Cys Ser Ser Ile Pro Asn Val Pro
          100             105             110
Gly Ser Cys Lys Glu Thr Phe Asn Leu Tyr Tyr Tyr Glu Ser Asp Phe
          115             120             125
Asp Ser Ala Thr Lys Thr Phe Pro Asn Trp Met Glu Asn Pro Trp Met
          130             135             140
Lys Val Asp Thr Ile Ala Ala Asp Glu Ser Phe Ser Gln Val Asp Leu
          145             150             155             160
Gly Gly Arg Val Met Lys Ile Asn Thr Glu Val Arg Ser Phe Gly Pro
          165             170             175
Val Ser Lys Asn Gly Phe Tyr Leu Ala Phe Gln Asp Tyr Gly Gly Cys
          180             185             190
Met Ser Leu Ile Ala Val Arg Val Phe Tyr Arg Lys Cys Pro Arg Val
          195             200             205
Ile Gln Asn Gly Ala Val Phe Gln Glu Thr Leu Ser Gly Ala Glu Ser
          210             215             220
Thr Ser Leu Val Ala Ala Arg Gly Thr Cys Ile Ser Asn Ala Glu Glu
          225             230             235             240
Val Asp Val Pro Ile Lys Leu Tyr Cys Asn Gly Asp Gly Glu Trp Leu
          245             250             255
Val Pro Ile Gly Arg Cys Met Cys Arg Pro Gly Tyr Glu Ser Val Glu
          260             265             270
Asn Gly Thr Val Cys Arg Gly Cys Pro Ser Gly Thr Phe Lys Ala Ser
          275             280             285

```

Gln Gly Asp Glu Gly Cys Val His Cys Pro Ile Asn Ser Arg Thr Thr
 290 295 300
 Ser Glu Gly Ala Thr Asn Cys Val Cys Arg Asn Gly Tyr Tyr Arg Ala
 305 310 315 320
 Asp Ala Asp Pro Val Asp Met Pro Cys Thr Thr Ile Pro Ser Ala Pro
 325 330 335
 Gln Ala Val Ile Ser Ser Val Asn Glu Thr Ser Leu Met Leu Glu Trp
 340 345 350
 Thr Pro Pro Arg Asp Ser Gly Gly Arg Glu Asp Leu Val Tyr Asn Ile
 355 360 365
 Ile Cys Lys Ser Cys Gly Ser Gly Arg Gly Ala Cys Thr Arg Cys Gly
 370 375 380
 Asp Asn Val Gln Phe Ala Pro Arg Gln Leu Gly Leu Thr Glu Pro Arg
 385 390 395 400
 Ile Tyr Ile Ser Asp Leu Leu Ala His Thr Gln Tyr Thr Phe Glu Ile
 405 410 415
 Gln Ala Val Asn Gly Val Thr Asp Gln Ser Pro Phe Ser Pro Gln Phe
 420 425 430
 Ala Ser Val Asn Ile Thr Thr Asn Gln Ala Ala Pro Ser Ala Val Ser
 435 440 445
 Ile Met His Gln Val Ser Arg Thr Val Asp Ser Ile Thr Leu Ser Trp
 450 455 460
 Ser Gln Pro Asp Gln Pro Asn Gly Val Ile Leu Asp Tyr Glu Leu Gln
 465 470 475 480
 Tyr Tyr Glu Lys Asn Leu Ser Glu Leu Asn Ser Thr Ala Val Lys Ser
 485 490 495
 Pro Thr Asn Thr Val Thr Val Gln Asn Leu Lys Ala Gly Thr Ile Tyr
 500 505 510
 Val Phe Gln Val Arg Ala Arg Thr Val Ala Gly Tyr Gly Arg Tyr Ser
 515 520 525
 Gly Lys Met Tyr Phe Gln Thr Met Thr Glu Ala Glu Tyr Gln Thr Ser
 530 535 540
 Val Gln Glu Lys Leu Pro Leu Ile Ile Gly Ser Ser Ala Ala Gly Leu
 545 550 555 560
 Val Phe Leu Ile Ala Val Val Val Ile Ile Ile Val Cys Asn Arg Arg
 565 570 575
 Arg Gly Phe Glu Arg Ala Asp Ser Glu Tyr Thr Asp Lys Leu Gln His
 580 585 590
 Tyr Thr Ser Gly His Met Thr Pro Gly Met Lys Ile Tyr Ile Asp Pro
 595 600 605
 Phe Thr Tyr Glu Asp Pro Asn Glu Ala Val Arg Glu Phe Ala Lys Glu
 610 615 620
 Ile Asp Ile Ser Cys Val Lys Ile Glu Gln Val Ile Gly Ala Gly Glu
 625 630 635 640

Phe Gly Glu Val Cys Ser Gly His Leu Lys Leu Pro Gly Lys Arg Glu
 645 650 655
 Ile Phe Val Ala Ile Lys Thr Leu Lys Ser Gly Tyr Thr Glu Lys Gln
 660 665 670
 Arg Arg Asp Phe Leu Ser Glu Ala Ser Ile Met Gly Gln Phe Asp His
 675 680 685
 Pro Asn Val Ile His Leu Glu Gly Val Val Thr Lys Ser Ser Pro Val
 690 695 700
 Met Ile Ile Thr Glu Phe Met Glu Asn Gly Ser Leu Asp Ser Phe Leu
 705 710 715 720
 Arg Gln Asn Asp Gly Gln Phe Thr Val Ile Gln Leu Val Gly Met Leu
 725 730 735
 Arg Gly Ile Ala Ala Gly Met Lys Tyr Leu Ala Asp Met Asn Tyr Val
 740 745 750
 His Arg Asp Leu Ala Ala Arg Asn Ile Leu Val Asn Ser Asn Leu Val
 755 760 765
 Cys Lys Val Ser Asp Phe Gly Leu Ser Arg Phe Leu Glu Asp Asp Thr
 770 775 780
 Ser Asp Pro Thr Tyr Thr Ser Ala Leu Gly Gly Lys Ile Pro Ile Arg
 785 790 795 800
 Trp Thr Ala Pro Glu Ala Ile Gln Tyr Arg Lys Phe Thr Ser Ala Ser
 805 810 815
 Asp Val Trp Ser Tyr Gly Ile Val Met Trp Glu Val Met Ser Tyr Gly
 820 825 830
 Glu Arg Pro Tyr Trp Asp Met Thr Asn Gln Asp Val Ile Asn Ala Ile
 835 840 845
 Glu Gln Asp Tyr Arg Leu Pro Pro Pro Met Asp Cys Pro Asn Ala Leu
 850 855 860
 His Gln Leu Met Leu Asp Cys Trp Gln Lys Asp Arg Asn His Arg Pro
 865 870 875 880
 Lys Phe Gly Gln Ile Val Asn Thr Leu Asp Lys Met Ile Arg Asn Pro
 885 890 895
 Asn Ser Leu Lys Ala Met Ala Pro Leu Ser Ser Gly Val Asn Leu Pro
 900 905 910
 Leu Leu Asp Arg Thr Ile Pro Asp Tyr Thr Ser Phe Asn Thr Val Asp
 915 920 925
 Glu Trp Leu Asp Ala Ile Lys Met Ser Gln Tyr Lys Glu Ser Phe Ala
 930 935 940
 Ser Ala Gly Phe Thr Thr Phe Asp Ile Val Ser Gln Met Thr Val Glu
 945 950 955 960
 Asp Ile Leu Arg Val Gly Val Thr Leu Ala Gly His Gln Lys Lys Ile
 965 970 975
 Leu Asn Ser Ile Gln Val Met Arg Ala Gln Met Asn Gln Ile Gln Ser
 980 985 990

100

Val Glu Val
995

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3125 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..2233

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

C CTC AAA TTC ACC CTG AGG GAC TGT AAC AGC CTT CCA GGA GGA CTT	46
Leu Lys Phe Thr Leu Arg Asp Cys Asn Ser Leu Pro Gly Gly Leu	
1 5 10 15	
GGG ACT TGC AAG GAG ACT TTT AAC ATG TAC TAC TTT GAG TCA GAT GAT	94
Gly Thr Cys Lys Glu Thr Phe Asn Met Tyr Tyr Phe Glu Ser Asp Asp	
20 25 30	
GAA GAT GGG AGG AAC ATC AGA GAG AAT CAG TAC ATC AAG ATA GAT ACC	142
Glu Asp Gly Arg Asn Ile Arg Glu Asn Gln Tyr Ile Lys Ile Asp Thr	
35 40 45	
ATT GCT GCT GAT GAG AGC TTC ACG GAG TTG GAC CTC GGC GAC AGA GTT	190
Ile Ala Ala Asp Glu Ser Phe Thr Glu Leu Asp Leu Gly Asp Arg Val	
50 55 60	
ATG AAG TTA AAC ACA GAA GTG AGA GAT GTT GGG CCT CTA ACA AAA AAA	238
Met Lys Leu Asn Thr Glu Val Arg Asp Val Gly Pro Leu Thr Lys Lys	
65 70 75	
GGG TTT TAC CTT GCT TTC CAG GAT GTG GGC GCC TGC ATT GCC CTG GTC	286
Gly Phe Tyr Leu Ala Phe Gln Asp Val Gly Ala Cys Ile Ala Leu Val	
80 85 90 95	
TCT GTG CGT GTG TAC TAC AAG AAA TGC CCA TCA GTG ATC CGC AAC CTG	334
Ser Val Arg Val Tyr Tyr Lys Lys Cys Pro Ser Val Ile Arg Asn Leu	
100 105 110	
GCA CGC TTT CCA GAT ACC ATC ACA GGA GCA GAT TCC TCG CAG CTG CTA	382
Ala Arg Phe Pro Asp Thr Ile Thr Gly Ala Asp Ser Ser Gln Leu Leu	
115 120 125	
GAA GTG TCA GGC GTC TGT GTC AAC CAC TCA GTG ACT GAT GAG GCA CCA	430
Glu Val Ser Gly Val Cys Val Asn His Ser Val Thr Asp Glu Ala Pro	
130 135 140	
AAG ATG CAC TGC AGT TCA GAG GGA GAA TGG CTG GTG CCC ATT GGG AAG	478
Lys Met His Cys Ser Ser Glu Gly Glu Trp Leu Val Pro Ile Gly Lys	
145 150 155	
TGT TTG TGC AAG GCA GGG TAC GAG GAG AAG AAC AAC ACC TGC CAA GCA	526
Cys Leu Cys Lys Ala Gly Tyr Glu Glu Lys Asn Asn Thr Cys Gln Ala	
160 165 170 175	
CCT TCT CCA GTC AGT AGT GTG AAA AAA GGG AAG ATA ACT AAA AAT AGC	574
Pro Ser Pro Val Ser Ser Val Lys Lys Gly Lys Ile Thr Lys Asn Ser	
180 185 190	

101

ATC TCC CTT TCC TGG CAG GAG CCA GAT CGA CCC AAC GGC ATC ATC CTG Ile Ser Leu Ser Trp Gln Glu Pro Asp Arg Pro Asn Gly Ile Ile Leu 195 200 205	622
GAA TAC GAA ATC AAA TAT TTT GAA AAG GAC CAG GAG ACA AGC TAC ACC Glu Tyr Glu Ile Lys Tyr Phe Glu Lys Asp Gln Glu Thr Ser Tyr Thr 210 215 220	670
ATC ATC AAA TCC AAA GAG ACC GCA ATT ACG GCA GAT GGC TTG AAA CCA Ile Ile Lys Ser Lys Glu Thr Ala Ile Thr Ala Asp Gly Leu Lys Pro 225 230 235	718
GGC TCA GCG TAC GTC TTC CAG ATC CGA GCC CGG ACA GCT GCT GGC TAC Gly Ser Ala Tyr Val Phe Gln Ile Arg Ala Arg Thr Ala Ala Gly Tyr 240 245 250 255	766
GGT GGC TTC AGT CGA AGA TTT GAG TTT GAA ACC AGC CCA GTG TTA GCT Gly Gly Phe Ser Arg Arg Phe Glu Phe Glu Thr Ser Pro Val Leu Ala 260 265 270	814
GCA TCC AGT GAC CAG AGC CAG ATT CCT ATA ATT GTT GTG TCT GTA ACA Ala Ser Ser Asp Gln Ser Gln Ile Pro Ile Ile Val Val Ser Val Thr 275 280 285	862
GTG GGA GTT ATT CTG CTG GCT GTT GTT ATC GGT TTC CTT CTC AGT GGA Val Gly Val Ile Leu Leu Ala Val Val Ile Gly Phe Leu Leu Ser Gly 290 295 300	910
AGT TGC TGC GAT CAT GGC TGT GGG TGG GCT TCT TCT CTG CGT GCT GTT Ser Cys Cys Asp His Gly Cys Gly Trp Ala Ser Ser Leu Arg Ala Val 305 310 315	958
GCC TAT CCG AGC CTA ATA TGG CGC TGT GGC TAC AGC AAG GCT AAA CAA Ala Tyr Pro Ser Leu Ile Trp Arg Cys Gly Tyr Ser Lys Ala Lys Gln 320 325 330 335	1006
GAC CCA GAA GAA GAA AAG ATG CAT TTT CAT AAT GGC CAC ATT AAA CTG Asp Pro Glu Glu Glu Lys Met His Phe His Asn Gly His Ile Lys Leu 340 345 350	1054
CCT GGT GTA AGA ACC TAC ATT GAT CCC CAC ACC TAT GAG GAC CCT AAT Pro Gly Val Arg Thr Tyr Ile Asp Pro His Thr Tyr Glu Asp Pro Asn 355 360 365	1102
CAA GCT GTC CAC GAG TTT GCC AAG GAA ATA GAA GCT TCG TGC ATA ACC Gln Ala Val His Glu Phe Ala Lys Glu Ile Glu Ala Ser Cys Ile Thr 370 375 380	1150
ATC GAG AGA GTT ATC GGA GCT GGT GAA TTT GGA GAA GTC TGC AGT GGA Ile Glu Arg Val Ile Gly Ala Gly Glu Phe Gly Glu Val Cys Ser Gly 385 390 395	1198
CGG CTG AAA CTG CAG GGA AAA CGC GAG TTT CCA GTG GCT ATC AAA ACC Arg Leu Lys Leu Gln Gly Lys Arg Glu Phe Pro Val Ala Ile Lys Thr 400 405 410 415	1246
CTG AAG GTG GGC TAC ACA GAG AAG CAA AGG CGA GAT TTC CTG GGA GAA Leu Lys Val Gly Tyr Thr Glu Lys Gln Arg Arg Asp Phe Leu Gly Glu 420 425 430	1294
GCG AGC ATC ATG GGG CAG TTC GAC CAC CCC AAC ATC ATC CAC CTG GAA Ala Ser Ile Met Gly Gln Phe Asp His Pro Asn Ile Ile His Leu Glu 435 440 445	1342
GGT GTC GTC ACA AAA AGC AAA CCT GTA ATG ATA GTA ACG GAA TAC ATG Gly Val Val Thr Lys Ser Lys Pro Val Met Ile Val Thr Glu Tyr Met 450 455 460	1390

102

GAA Glu 465	AAT Asn 465	GGT Gly 465	TCT Ser 465	CTG Leu 465	GAT Asp 465	ACA Thr 465	TTT Phe 465	TTA Leu 465	AAG Lys 465	AAG Lys 465	AAC Asn 465	GAT Asp 465	GGG Gly 465	CAG Gln 465	TTC Phe 465	1438
ACG Thr 480	GTC Val 480	ATT Ile 480	CAG Gln 480	CTG Leu 480	GTC Val 480	GGG Gly 480	ATG Met 480	CTG Val 480	CGA Arg 480	GGC Gly 480	ATC Ile 480	GCA Ala 480	TCA Ser 480	GGG Gly 480	ATG Met 480	1486
AAG Lys 500	TAC Tyr 500	CTG Leu 500	TCT Ser 500	GAC Asp 500	ATG Met 500	GGT Gly 500	TAC Tyr 500	GTA Val 500	CAC His 500	AGA Arg 500	GAC Asp 500	CTC Leu 500	GCT Ala 500	GCC Ala 500	AGG Arg 500	1534
AAT Asn 515	ATC Ile 515	CTC Leu 515	ATC Ile 515	AAC Asn 515	AGC Ser 515	AAC Asn 515	TTA Leu 515	GTC Val 515	TGC Cys 515	AAG Lys 515	GTG Val 515	TCT Ser 515	GAC Asp 515	TTT Phe 515	GGC Gly 515	1582
CTC Leu 530	TCC Ser 530	AGA Val 530	GTC Val 530	CTA Leu 530	GAA Glu 530	GAT Asp 530	GAT Pro 530	CCT Pro 530	GAA Glu 530	GCA Ala 530	GCG Ala 530	TAC Tyr 530	ACA Thr 530	ACC Thr 530	AGG Arg 530	1630
GGA Gly 545	GGG Gly 545	AAG Lys 545	ATC Ile 545	CCC Pro 545	ATC Ile 545	CGA Arg 545	TGG Trp 545	ACG Thr 545	GCA Ala 545	CCT Pro 545	GAA Glu 545	GCA Ala 545	ATC Ile 545	GCC Ala 545	TTC Phe 545	1678
CGC Arg 560	AAA Lys 560	TTC Phe 560	ACG Thr 560	TCG Ser 560	GCC Ala 560	AGC Ser 560	GAT Asp 560	GTG Val 560	TGG Trp 560	AGC Ser 560	TAC Tyr 560	GGC Gly 560	ATT Ile 560	GTG Val 560	ATG Met 560	1726
TGG Trp 580	GAA Glu 580	GTG Val 580	ATG Met 580	TCC Ser 580	TAT Tyr 580	GGC Gly 580	GAG Glu 580	AGA Arg 580	CCT Pro 580	TAC Tyr 580	TGG Trp 580	GAA Glu 580	ATG Met 580	ACA Thr 580	AAC Asn 580	1774
CAA Gln 595	GAT Asp 595	GTG Val 595	ATT Ile 595	AAA Lys 595	GCC Ala 595	GTG Val 595	GAG Glu 595	GAA Gly 595	GGC Tyr 595	TAT Arg 595	CGC Leu 595	CTG Pro 595	CCA Ser 595	AGT Ser 595	CCC Pro 595	1822
ATG Met 610	GAC Asp 610	TGC Cys 610	CCT Pro 610	GCT Ala 610	GCT Ala 610	CTC Leu 610	TAC Tyr 610	CAG Gln 610	TTG Leu 610	ATG Met 610	CTT Leu 610	GAC Asp 610	TGC Cys 610	TGG Trp 610	CAG Gln 610	1870
AAA Lys 625	GAC Asp 625	CGC Arg 625	AAC Asn 625	AGC Ser 625	AGG Arg 625	CCC Pro 625	AAG Lys 625	TTT Phe 625	GAT Asp 625	GAA Glu 625	ATT Ile 625	GTC Val 625	AGC Ser 625	ATG Met 625	TTG Leu 625	1918
GAC Asp 640	AAG Lys 640	CTC Leu 640	ATC Ile 640	CGT Arg 640	AAC Pro 640	CCA Ser 640	AGC Ser 640	AGC Ser 640	TTG Leu 640	AAG Lys 640	ACG Thr 640	TTG Leu 640	GTT Val 640	AAT Asn 640	GCA Ala 640	1966
TCG Ser 660	AGC Ser 660	AGA Arg 660	GTA Val 660	TCA Ser 660	AAT Asn 660	TTG Leu 660	TTG Leu 660	GTA Val 660	GAA Glu 660	CAC His 660	AGT Ser 660	CCA Pro 660	GTG Val 660	GGG Gly 660	AGC Ser 660	2014
GGT Gly 675	GCC Ala 675	TAC Tyr 675	AGG Arg 675	TCA Ser 675	GTG Val 675	GGT Gly 675	GAG Glu 675	TGG Trp 675	CTG Leu 675	GAA Glu 675	GCC Ala 675	ATC Ile 675	AAA Lys 675	ATG Met 675	GGT Gly 675	2062
CGA Arg 690	TAC Tyr 690	ACC Glu 690	GAG Ile 690	ATT Phe 690	TTC Met 690	ATG Met 690	GAG Glu 690	AAT Asn 690	GGA Gly 690	TAC Tyr 690	AGT Ser 690	TCG Ser 690	ATG Met 690	GAT Asp 690	TCT Ser 690	2110
GTG Val 705	GCT Ala 705	CAG Gln 705	GTG Val 705	ACC Thr 705	CTA Leu 705	GAG Glu 705	GAT Asp 705	TTG Leu 705	AGG Arg 705	CGG Arg 705	CTG Leu 705	GGA Gly 705	GTG Val 705	ACA Thr 705	CTT Leu 705	2158
GTT Val 720	GGT Gly 720	CAC His 720	CAG Gln 720	AAG Lys 720	AAG Lys 720	ATA Ile 720	ATG Met 720	AAC Asn 720	AGC Ser 720	CTT Leu 720	CAA Gln 720	GAG Glu 720	ATG Met 720	AAG Lys 720	GTC Val 720	2206

103

CAG TTG GTG AAT GGG ATG GTG CCA TTG TAACTCGGTT TTTAAGTCAC 2253
 Gln Leu Val Asn Gly Met Val Pro Leu
 740

TTCTCTGAGT GGTCGGTCCT GCACTTTGTA TACTAGCTCT GAGATTTATT TTGACTAAAG 2313
 AAGAAAAAAG GGAAATTCAG TGGTTTCTGT AACTGAAGGA CGCTGGCTTC TGCCACAGCA 2373
 TTTATAAAGC AGTGTTTGAC TGAAGTTTTTCT ATTTTCTTCC TATTGTGTGC CTCATTCTCA 2433
 TGAAGTAAAT GTAACATGCA TGGAACATGG AAATGGATCT ACTGTACATG AGGTTACCCA 2493
 ATTTCTTGCG CTTTCAGCATG ACAACAGCAA GCCTTCCCAC CACATGTTGT CTATACATGG 2553
 GAGATATATA TATATGCATA TATATATATA GCACCTTTAT ATACTGAAAT ACAGCAGCAG 2613
 CACATGTTAA TACTTCCAAG GACTTACTTG ACTAGAGAAG TTTTGCAGCC ATTGTGGGCT 2673
 CACACAAGCT GCGGTTTACT GAAGTTTACT TCAAGTCTTA CTTGTCTACA GAAGTGTTAT 2733
 GAAGAGCAAT ATGATTAGAT TATTTCTGGA TAGATATTTT GTTTTGTAAG TTTAAAAAAT 2793
 CGTGTACAC AGCGTTAAGT TATAGAGACT AGTGTATAAA CATGTTGCTT GCTCAATGGC 2853
 AAATACAATA CAGGGTGTAT ATTTTCTTCT CTCTGTGTTG CAAAGTTCTT TTAGTTTGCT 2913
 CTTCTGTGAG GATAATACGT TATGATGTAT ATACTGTACA GTTTGCTACA CATCAGGTAC 2973
 AAGATTGGGG CTTTCTCAAT GTTTTGTCTT TTTCCCTCT TTTGTTTCAT TTTGTCTTCC 3033
 TTTTGTGTTA ACCACTATGC TTTGTATTTT TGCTGCTGTT TGGTTTGAGG CAACATATAA 3093
 AGCTTTCAGG TGTTTTGATT ATAAAAAAA AG 3125

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 744 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Leu Lys Phe Thr Leu Arg Asp Cys Asn Ser Leu Pro Gly Gly Leu Gly
 1 5 10 15

Thr Cys Lys Glu Thr Phe Asn Met Tyr Tyr Phe Glu Ser Asp Asp Glu
 20 25 30

Asp Gly Arg Asn Ile Arg Glu Asn Gln Tyr Ile Lys Ile Asp Thr Ile
 35 40 45

Ala Ala Asp Glu Ser Phe Thr Glu Leu Asp Leu Gly Asp Arg Val Met
 50 55 60

Lys Leu Asn Thr Glu Val Arg Asp Val Gly Pro Leu Thr Lys Lys Gly
 65 70 75 80

Phe Tyr Leu Ala Phe Gln Asp Val Gly Ala Cys Ile Ala Leu Val Ser
 85 90 95

Val Arg Val Tyr Tyr Lys Lys Cys Pro Ser Val Ile Arg Asn Leu Ala
 100 105 110

104

Arg Phe Pro Asp Thr Ile Thr Gly Ala Asp Ser Ser Gln Leu Leu Glu
 115 120 125
 Val Ser Gly Val Cys Val Asn His Ser Val Thr Asp Glu Ala Pro Lys
 130 135 140
 Met His Cys Ser Ser Glu Gly Glu Trp Leu Val Pro Ile Gly Lys Cys
 145 150 155 160
 Leu Cys Lys Ala Gly Tyr Glu Glu Lys Asn Asn Thr Cys Gln Ala Pro
 165 170 175
 Ser Pro Val Ser Ser Val Lys Lys Gly Lys Ile Thr Lys Asn Ser Ile
 180 185 190
 Ser Leu Ser Trp Gln Glu Pro Asp Arg Pro Asn Gly Ile Ile Leu Glu
 195 200 205
 Tyr Glu Ile Lys Tyr Phe Glu Lys Asp Gln Glu Thr Ser Tyr Thr Ile
 210 215 220
 Ile Lys Ser Lys Glu Thr Ala Ile Thr Ala Asp Gly Leu Lys Pro Gly
 225 230 235 240
 Ser Ala Tyr Val Phe Gln Ile Arg Ala Arg Thr Ala Ala Gly Tyr Gly
 245 250 255
 Gly Phe Ser Arg Arg Phe Glu Phe Glu Thr Ser Pro Val Leu Ala Ala
 260 265 270
 Ser Ser Asp Gln Ser Gln Ile Pro Ile Ile Val Val Ser Val Thr Val
 275 280 285
 Gly Val Ile Leu Leu Ala Val Val Ile Gly Phe Leu Leu Ser Gly Ser
 290 295 300
 Cys Cys Asp His Gly Cys Gly Trp Ala Ser Ser Leu Arg Ala Val Ala
 305 310 315 320
 Tyr Pro Ser Leu Ile Trp Arg Cys Gly Tyr Ser Lys Ala Lys Gln Asp
 325 330 335
 Pro Glu Glu Glu Lys Met His Phe His Asn Gly His Ile Lys Leu Pro
 340 345 350
 Gly Val Arg Thr Tyr Ile Asp Pro His Thr Tyr Glu Asp Pro Asn Gln
 355 360 365
 Ala Val His Glu Phe Ala Lys Glu Ile Glu Ala Ser Cys Ile Thr Ile
 370 375 380
 Glu Arg Val Ile Gly Ala Gly Glu Phe Gly Glu Val Cys Ser Gly Arg
 385 390 395 400
 Leu Lys Leu Gln Gly Lys Arg Glu Phe Pro Val Ala Ile Lys Thr Leu
 405 410 415
 Lys Val Gly Tyr Thr Glu Lys Gln Arg Arg Asp Phe Leu Gly Glu Ala
 420 425 430
 Ser Ile Met Gly Gln Phe Asp His Pro Asn Ile Ile His Leu Glu Gly
 435 440 445
 Val Val Thr Lys Ser Lys Pro Val Met Ile Val Thr Glu Tyr Met Glu
 450 455 460

105

Asn Gly Ser Leu Asp Thr Phe Leu Lys Lys Asn Asp Gly Gln Phe Thr
 465 470 475 480
 Val Ile Gln Leu Val Gly Met Leu Arg Gly Ile Ala Ser Gly Met Lys
 485 490 495
 Tyr Leu Ser Asp Met Gly Tyr Val His Arg Asp Leu Ala Ala Arg Asn
 500 505 510
 Ile Leu Ile Asn Ser Asn Leu Val Cys Lys Val Ser Asp Phe Gly Leu
 515 520 525
 Ser Arg Val Leu Glu Asp Asp Pro Glu Ala Ala Tyr Thr Thr Arg Gly
 530 535 540
 Gly Lys Ile Pro Ile Arg Trp Thr Ala Pro Glu Ala Ile Ala Phe Arg
 545 550 555 560
 Lys Phe Thr Ser Ala Ser Asp Val Trp Ser Tyr Gly Ile Val Met Trp
 565 570 575
 Glu Val Met Ser Tyr Gly Glu Arg Pro Tyr Trp Glu Met Thr Asn Gln
 580 585 590
 Asp Val Ile Lys Ala Val Glu Glu Gly Tyr Arg Leu Pro Ser Pro Met
 595 600 605
 Asp Cys Pro Ala Ala Leu Tyr Gln Leu Met Leu Asp Cys Trp Gln Lys
 610 615 620
 Asp Arg Asn Ser Arg Pro Lys Phe Asp Glu Ile Val Ser Met Leu Asp
 625 630 635 640
 Lys Leu Ile Arg Asn Pro Ser Ser Leu Lys Thr Leu Val Asn Ala Ser
 645 650 655
 Ser Arg Val Ser Asn Leu Leu Val Glu His Ser Pro Val Gly Ser Gly
 660 665 670
 Ala Tyr Arg Ser Val Gly Glu Trp Leu Glu Ala Ile Lys Met Gly Arg
 675 680 685
 Tyr Thr Glu Ile Phe Met Glu Asn Gly Tyr Ser Ser Met Asp Ser Val
 690 695 700
 Ala Gln Val Thr Leu Glu Asp Leu Arg Arg Leu Gly Val Thr Leu Val
 705 710 715 720
 Gly His Gln Lys Lys Ile Met Asn Ser Leu Gln Glu Met Lys Val Gln
 725 730 735
 Leu Val Asn Gly Met Val Pro Leu
 740

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3056 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

106

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 2..2131

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

C CTC AAA TTC ACC CTG AGG GAC TGT AAC AGC CTT CCA GGA GGA CTT	46
Leu Lys Phe Thr Leu Arg Asp Cys Asn Ser Leu Pro Gly Gly Leu	
1 5 10 15	
GGG ACT TGC AAG GAG ACT TTT AAC ATG TAC TAC TTT GAG TCA GAT GAT	94
Gly Thr Cys Lys Glu Thr Phe Asn Met Tyr Tyr Phe Glu Ser Asp Asp	
20 25 30	
GAA GAT GGG AGG AAC ATC AGA GAG AAT CAG TAC ATC AAG ATA GAT ACC	142
Glu Asp Gly Arg Asn Ile Arg Glu Asn Gln Tyr Ile Lys Ile Asp Thr	
35 40 45	
ATT GCT GCT GAT GAG AGC TTC ACG GAG TTG GAC CTC GGC GAC AGA GTT	190
Ile Ala Ala Asp Glu Ser Phe Thr Glu Leu Asp Leu Gly Asp Arg Val	
50 55 60	
ATG AAG TTA AAC ACA GAA GTG AGA GAT GTT GGG CCT CTA ACA AAA AAA	238
Met Lys Leu Asn Thr Glu Val Arg Asp Val Gly Pro Leu Thr Lys Lys	
65 70 75	
GGA TTT TAC CTT GCT TTC CAG GAT GTG GGC GCC TGC ATT GCC CTG GTC	286
Gly Phe Tyr Leu Ala Phe Gln Asp Val Gly Ala Cys Ile Ala Leu Val	
80 85 90 95	
TCT GTG CGT GTG TAC TAC AAG AAA TGC CCA TCA GTG ATC CGC AAC CTG	334
Ser Val Arg Val Tyr Tyr Lys Lys Cys Pro Ser Val Ile Arg Asn Leu	
100 105 110	
GCA CGC TTT CCA GAT ACC ATC ACA GGA GCA GAT TCC TCG CAG CTG CTA	382
Ala Arg Phe Pro Asp Thr Ile Thr Gly Ala Asp Ser Ser Gln Leu Leu	
115 120 125	
GAA GTG TCA GGC GTC TGT GTC AAC CAC TCA GTG ACT GAT GAG GCA CCA	430
Glu Val Ser Gly Val Cys Val Asn His Ser Val Thr Asp Glu Ala Pro	
130 135 140	
AAG ATG CAC TGC AGT TCA GAG GGA GAA TGG CTG GTG CCC ATT GGG AAG	478
Lys Met His Cys Ser Ser Glu Gly Glu Trp Leu Val Pro Ile Gly Lys	
145 150 155	
TGT TTG TGC AAG GCA GGG TAC GAG GAG AAG AAC AAC ACC TGC CAA GCA	526
Cys Leu Cys Lys Ala Gly Tyr Glu Glu Lys Asn Asn Thr Cys Gln Ala	
160 165 170 175	
CCT TCT CCA GTC AGT AGT GTG AAA AAA GGG AAG ATA ACT AAA AAT AGC	574
Pro Ser Pro Val Ser Ser Val Lys Lys Gly Lys Ile Thr Lys Asn Ser	
180 185 190	
ATC TCC CTT TCC TGG CAG GAG CCA GAT CGA CCC AAC GGC ATC ATC CTG	622
Ile Ser Leu Ser Trp Gln Glu Pro Asp Arg Pro Asn Gly Ile Ile Leu	
195 200 205	
GAA TAC GAA ATC AAA TAT TTT GAA AAG GAC CAG GAG ACA AGC TAC ACC	670
Glu Tyr Glu Ile Lys Tyr Phe Glu Lys Asp Gln Glu Thr Ser Tyr Thr	
210 215 220	
ATC ATC AAA TCC AAA GAG ACC GCA ATT ACG GCA GAT GGC TTG AAA CCA	718
Ile Ile Lys Ser Lys Glu Thr Ala Ile Thr Ala Asp Gly Leu Lys Pro	
225 230 235	

107

GGC TCA GCG TAC GTC TTC CAG ATC CGA GCC CGG ACA GCT GCT GGC TAC Gly Ser Ala Tyr Val Phe Gln Ile Arg Ala Arg Thr Ala Ala Gly Tyr 240 245 250 255	766
GGT GGC TTC AGT CGA AGA TTT GAG TTT GAA ACC AGC CCA GTG TTA GCT Gly Gly Phe Ser Arg Arg Phe Glu Phe Glu Thr Ser Pro Val Leu Ala 260 265 270	814
GCA TCC AGT GAC CAG AGC CAG ATT CCT ATA ATT GTT GTG TCT GTA ACA Ala Ser Ser Asp Gln Ser Gln Ile Pro Ile Ile Val Val Ser Val Thr 275 280 285	862
GTG GGA GTT ATT CTG CTG GCT GTT GTT ATC GGT TTC CTT CTC AGT GGA Val Gly Val Ile Leu Leu Ala Val Val Ile Gly Phe Leu Leu Ser Gly 290 295 300	910
AGG CGC TGT GGC TAC AGC AAG GCT AAA CAA GAC CCA GAA GAA GAA AAG Arg Arg Cys Gly Tyr Ser Lys Ala Lys Gln Asp Pro Glu Glu Glu Lys 305 310 315	958
ATG CAT TTT CAT AAT GGC CAC ATT AAA CTG CCT GGT GTA AGA ACC TAC Met His Phe His Asn Gly His Ile Lys Leu Pro Gly Val Arg Thr Tyr 320 325 330 335	1006
ATT GAT CCC CAC ACC TAT GAG GAC CCT AAT CAA GCT GTC CAC GAG TTT Ile Asp Pro His Thr Tyr Glu Asp Pro Asn Gln Ala Val His Glu Phe 340 345 350	1054
GCC AAG GAA ATA GAA GCT TCG TGC ATA ACC ATC GAG AGA GTT ATC GGA Ala Lys Glu Ile Glu Ala Ser Cys Ile Thr Ile Glu Arg Val Ile Gly 355 360 365	1102
GCT GGT GAA TTT GGA GAA GTC TGC AGT GGA CGG CTG AAA CTG CAG GGA Ala Gly Glu Phe Gly Glu Val Cys Ser Gly Arg Leu Lys Leu Gln Gly 370 375 380	1150
AAA CGC GAG TTT CCA GTG GCT ATC AAA ACC CTG AAG GTG GGC TAC ACA Lys Arg Glu Phe Pro Val Ala Ile Lys Thr Leu Lys Val Gly Tyr Thr 385 390 395	1198
GAG AAG CAA AGG CGA GAT TTC CTG GGA GAA GCG AGC ATC ATG GGG CAG Glu Lys Gln Arg Arg Asp Phe Leu Gly Glu Ala Ser Ile Met Gly Gln 400 405 410 415	1246
TTC GAC CAC CCC AAC ATC ATC CAC CTG GAA GGT GTC GTC ACA AAA AGC Phe Asp His Pro Asn Ile Ile His Leu Glu Gly Val Val Thr Lys Ser 420 425 430	1294
AAA CCT GTA ATG ATA GTA ACG GAA TAC ATG GAA AAT GGT TCT CTG GAT Lys Pro Val Met Ile Val Thr Glu Tyr Met Glu Asn Gly Ser Leu Asp 435 440 445	1342
ACA TTT TTA AAG AAG AAC GAT GGG CAG TTC ACG GTC ATT CAG CTG GTC Thr Phe Leu Lys Lys Asn Asp Gly Gln Phe Thr Val Ile Gln Leu Val 450 455 460	1390
GGG ATG CTG CGA GGC ATC GCA TCA GGG ATG AAG TAC CTG TCT GAC ATG Gly Met Leu Arg Gly Ile Ala Ser Gly Met Lys Tyr Leu Ser Asp Met 465 470 475	1438
GGT TAC GTA CAC AGA GAC CTC GCT GCC AGG AAT ATC CTC ATC AAC AGC Gly Tyr Val His Arg Asp Leu Ala Ala Arg Asn Ile Leu Ile Asn Ser 480 485 490 495	1486
AAC TTA GTC TGC AAG GTG TCT GAC TTT GGC CTC TCC AGA GTC CTA GAA Asn Leu Val Cys Lys Val Ser Asp Phe Gly Leu Ser Arg Val Leu Glu 500 505 510	1534

108

GAT GAT CCT GAA GCA GCG TAC ACA ACC AGG GGA GGG AAG ATC CCC ATC Asp Asp Pro Glu Ala Ala Tyr Thr Thr Arg Gly Gly Lys Ile Pro Ile 515 520 525	1582
CGA TGG ACG GCA CCT GAA GCA ATC GCC TTC CGC AAA TTC ACG TCG GCC Arg Trp Thr Ala Pro Glu Ala Ile Ala Phe Arg Lys Phe Thr Ser Ala 530 535 540	1630
AGC GAT GTG TGG AGC TAC GGC ATT GTG ATG TGG GAA GTG ATG TCC TAT Ser Asp Val Trp Ser Tyr Gly Ile Val Met Trp Glu Val Met Ser Tyr 545 550 555	1678
GGC GAG AGA CCT TAC TGG GAA ATG ACA AAC CAA GAT GTG ATT AAA GCC Gly Glu Arg Pro Tyr Trp Glu Met Thr Asn Gln Asp Val Ile Lys Ala 560 565 570 575	1726
GTG GAG GAA GGC TAT CGC CTG CCA AGT CCC ATG GAC TGC CCT GCT GCT Val Glu Glu Gly Tyr Arg Leu Pro Ser Pro Met Asp Cys Pro Ala Ala 580 585 590	1774
CTC TAC CAG TTG ATG CTT GAC TGC TGG CAG AAA GAC CGC AAC AGC AGG Leu Tyr Gln Leu Met Leu Asp Cys Trp Gln Lys Asp Arg Asn Ser Arg 595 600 605	1822
CCC AAG TTT GAT GAA ATT GTC AGC ATG TTG GAC AAG CTC ATC CGT AAC Pro Lys Phe Asp Glu Ile Val Ser Met Leu Asp Lys Leu Ile Arg Asn 610 615 620	1870
CCA AGC AGC TTG AAG ACG TTG GTT AAT GCA TCG AGC AGA GTA TCA AAT Pro Ser Ser Leu Lys Thr Leu Val Asn Ala Ser Ser Arg Val Ser Asn 625 630 635	1918
TTG TTG GTA GAA CAC AGT CCA GTG GGG AGC GGT GCC TAC AGG TCA GTG Leu Leu Val Glu His Ser Pro Val Gly Ser Gly Ala Tyr Arg Ser Val 640 645 650 655	1966
GGT GAG TGG CTG GAA GCC ATC AAA ATG GGT CGA TAC ACC GAG ATT TTC Gly Glu Trp Leu Glu Ala Ile Lys Met Gly Arg Tyr Thr Glu Ile Phe 660 665 670	2014
ATG GAG AAT GGA TAC AGT TCG ATG GAT TCT GTG GCT CAG GTG ACC CTA Met Glu Asn Gly Tyr Ser Ser Met Asp Ser Val Ala Gln Val Thr Leu 675 680 685	2062
GAG GAC GAA TCA CCT TGT GAA AAG TGG AGC CTC ACC CTC CAC CCC CTC Glu Asp Glu Ser Pro Cys Glu Lys Trp Ser Leu Thr Leu His Pro Leu 690 695 700	2110
TTT CCA ACT GGA TAT CAG ACT TGAAGGAAAC CTTTCCAGTG GACCAGACCT Phe Pro Thr Gly Tyr Gln Thr 705 710	2161
GCTCTTTAAA CTGTGGACC ACCTAGTGAC TTTGAGTGTG TCTGGAGCTC TTTCAATCCA	2221
CTGCAAGAAT AACTTTACCA GGACAGTACT CAAGAATAGA TAGATCCATG ACATGAGTTT	2281
CAGTCTGATA TTGACTGGA CCAATTACTA ACAAATGTG GACTGCATAC TTACACCTTT	2341
TGAAAGATCT GTACTCACCG AATCTCAGGA CACCCTGTTG TTTGTTATTA GATGAAGAAC	2401
TCTGAATATT TGTAATAATA TGTGATGTGT TGCTTTGCAT TGTATTTTTT TCTTATAAAA	2461
TAAAATAAAT TATTTATTAA AAGTTATACT GGGATGAAGA CCATTTAAGA GTTCACCTGC	2521
TCTAGATGCT TATTTCTTAAC CTGAAACCTC AGTTCCGGAT AGTGATACTG CACACGCTTG	2581
TGAACAAACC CATTCTCGTG TCATAACCAA ACAGGATGGG AGTAATGAAT AAGAGCAGAT	2641

109

GAACTCTTAA AAGAAAGATC CTAATCTCAT GCAAAGGTCC CTTGCAAGTG GATTCCTCTC 2701
 TCCCTAGCGT CTTCTAAAGG TCTTTGAGGT TATTCTTTCC CCTCTTTCAA ACTGACAGCT 2761
 AACTCTGTGA GTAGTGTGAG TCTGCATGGG CCAGTGTAGA ACTGCACCAT GTTGAAGAAG 2821
 AGTGCTGCAA TATGGCTGGG GTGGGAGATG AAATGCAAAG TAATCTCTGG TAGGCTGATG 2881
 GCTTCCAGCC ATGGAGGTAT TTCAGGAACC TGGCCCTTTT GCTTGCATGA GTAATGAATG 2941
 GAGTGGTGAG GAGTGTGTGA TTTTATGTGG CAATCCAGTC CTAGTCTACA CTGTGTTTGA 3001
 CAAATTGGTC CATGGTGTAT AAGTAGTTCT ATTTGTAAAT AAAATGTTTT AAATG 3056

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 710 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Leu Lys Phe Thr Leu Arg Asp Cys Asn Ser Leu Pro Gly Gly Leu Gly
 1 5 10 15
 Thr Cys Lys Glu Thr Phe Asn Met Tyr Tyr Phe Glu Ser Asp Asp Glu
 20 25 30
 Asp Gly Arg Asn Ile Arg Glu Asn Gln Tyr Ile Lys Ile Asp Thr Ile
 35 40 45
 Ala Ala Asp Glu Ser Phe Thr Glu Leu Asp Leu Gly Asp Arg Val Met
 50 55 60
 Lys Leu Asn Thr Glu Val Arg Asp Val Gly Pro Leu Thr Lys Lys Gly
 65 70 75 80
 Phe Tyr Leu Ala Phe Gln Asp Val Gly Ala Cys Ile Ala Leu Val Ser
 85 90 95
 Val Arg Val Tyr Tyr Lys Lys Cys Pro Ser Val Ile Arg Asn Leu Ala
 100 105 110
 Arg Phe Pro Asp Thr Ile Thr Gly Ala Asp Ser Ser Gln Leu Leu Glu
 115 120 125
 Val Ser Gly Val Cys Val Asn His Ser Val Thr Asp Glu Ala Pro Lys
 130 135 140
 Met His Cys Ser Ser Glu Gly Glu Trp Leu Val Pro Ile Gly Lys Cys
 145 150 155 160
 Leu Cys Lys Ala Gly Tyr Glu Glu Lys Asn Asn Thr Cys Gln Ala Pro
 165 170 175
 Ser Pro Val Ser Ser Val Lys Lys Gly Lys Ile Thr Lys Asn Ser Ile
 180 185 190
 Ser Leu Ser Trp Gln Glu Pro Asp Arg Pro Asn Gly Ile Ile Leu Glu
 195 200 205
 Tyr Glu Ile Lys Tyr Phe Glu Lys Asp Gln Glu Thr Ser Tyr Thr Ile
 210 215 220

110

Ile Lys Ser Lys Glu Thr Ala Ile Thr Ala Asp Gly Leu Lys Pro Gly
 225 230 235 240
 Ser Ala Tyr Val Phe Gln Ile Arg Ala Arg Thr Ala Ala Gly Tyr Gly
 245 250 255
 Gly Phe Ser Arg Arg Phe Glu Phe Glu Thr Ser Pro Val Leu Ala Ala
 260 265 270
 Ser Ser Asp Gln Ser Gln Ile Pro Ile Ile Val Val Ser Val Thr Val
 275 280 285
 Gly Val Ile Leu Leu Ala Val Val Ile Gly Phe Leu Leu Ser Gly Arg
 290 295 300
 Arg Cys Gly Tyr Ser Lys Ala Lys Gln Asp Pro Glu Glu Glu Lys Met
 305 310 315 320
 His Phe His Asn Gly His Ile Lys Leu Pro Gly Val Arg Thr Tyr Ile
 325 330 335
 Asp Pro His Thr Tyr Glu Asp Pro Asn Gln Ala Val His Glu Phe Ala
 340 345 350
 Lys Glu Ile Glu Ala Ser Cys Ile Thr Ile Glu Arg Val Ile Gly Ala
 355 360 365
 Gly Glu Phe Gly Glu Val Cys Ser Gly Arg Leu Lys Leu Gln Gly Lys
 370 375 380
 Arg Glu Phe Pro Val Ala Ile Lys Thr Leu Lys Val Gly Tyr Thr Glu
 385 390 395 400
 Lys Gln Arg Arg Asp Phe Leu Gly Glu Ala Ser Ile Met Gly Gln Phe
 405 410 415
 Asp His Pro Asn Ile Ile His Leu Glu Gly Val Val Thr Lys Ser Lys
 420 425 430
 Pro Val Met Ile Val Thr Glu Tyr Met Glu Asn Gly Ser Leu Asp Thr
 435 440 445
 Phe Leu Lys Lys Asn Asp Gly Gln Phe Thr Val Ile Gln Leu Val Gly
 450 455 460
 Met Leu Arg Gly Ile Ala Ser Gly Met Lys Tyr Leu Ser Asp Met Gly
 465 470 475 480
 Tyr Val His Arg Asp Leu Ala Ala Arg Asn Ile Leu Ile Asn Ser Asn
 485 490 495
 Leu Val Cys Lys Val Ser Asp Phe Gly Leu Ser Arg Val Leu Glu Asp
 500 505 510
 Asp Pro Glu Ala Ala Tyr Thr Thr Arg Gly Gly Lys Ile Pro Ile Arg
 515 520 525
 Trp Thr Ala Pro Glu Ala Ile Ala Phe Arg Lys Phe Thr Ser Ala Ser
 530 535 540
 Asp Val Trp Ser Tyr Gly Ile Val Met Trp Glu Val Met Ser Tyr Gly
 545 550 555 560
 Glu Arg Pro Tyr Trp Glu Met Thr Asn Gln Asp Val Ile Lys Ala Val
 565 570 575

111

Glu Glu Gly Tyr Arg Leu Pro Ser Pro Met Asp Cys Pro Ala Ala Leu
 580 585 590
 Tyr Gln Leu Met Leu Asp Cys Trp Gln Lys Asp Arg Asn Ser Arg Pro
 595 600 605
 Lys Phe Asp Glu Ile Val Ser Met Leu Asp Lys Leu Ile Arg Asn Pro
 610 615 620
 Ser Ser Leu Lys Thr Leu Val Asn Ala Ser Ser Arg Val Ser Asn Leu
 625 630 635 640
 Leu Val Glu His Ser Pro Val Gly Ser Gly Ala Tyr Arg Ser Val Gly
 645 650 655
 Glu Trp Leu Glu Ala Ile Lys Met Gly Arg Tyr Thr Glu Ile Phe Met
 660 665 670
 Glu Asn Gly Tyr Ser Ser Met Asp Ser Val Ala Gln Val Thr Leu Glu
 675 680 685
 Asp Glu Ser Pro Cys Glu Lys Trp Ser Leu Thr Leu His Pro Leu Phe
 690 695 700
 Pro Thr Gly Tyr Gln Thr
 705 710

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 19 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Arg Ile Cys Thr Pro Asp Val Ser Gly Thr Val Gly Ser Arg Pro Ala
 1 5 10 15
 Ala Asp His

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Cys Leu Glu Thr His Thr Lys Asn Ser Pro Val Pro Val
 1 5 10

112

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Lys	Met	Gln	Gln	Met	His	Gly	Arg	Met	Val	Pro	Val
1				5					10		

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Lys	Val	His	Leu	Asn	Gln	Leu	Glu	Pro	Val	Glu	Val
1				5					10		

What is claimed is:

1. A composition of matter, comprising an isolated nucleic acid sequence encoding a Eph-related protein tyrosine kinase, or functional fragment thereof, having about 23 to 66 percent amino acid sequence identity
5 in its carboxyl terminal variable region compared to known members of the Eph subclass of tyrosine kinases.
2. The composition of claim 1, comprising substantially the same nucleotide sequence selected from the group consisting of SEQ ID NOS: 3, 7, 9, 11, 13, 19 and
10 21.
3. A composition of matter, comprising a vector containing the nucleic acid of claim 1.
4. The composition of claim 3, wherein said vector is for the expression of a recombinant Eph-related
15 protein tyrosine kinase.
5. The composition of claim 4, wherein said expression is in a procaryotic host.
6. The composition of claim 4, wherein said expression is in a eucaryotic host.
- 20 7. A composition of matter, comprising a host cell containing the vector of claim 3.
8. The composition of claim 7, wherein said host cell is procaryotic.
9. The composition of claim 7, wherein said
25 host cell is eucaryotic.

10. A composition of matter, comprising a substantially purified Eph-related protein tyrosine kinase, or functional fragment thereof, having about 23 to 66 percent amino acid sequence identity in its carboxyl terminal variable region compared to known members of the Eph subclass of tyrosine kinases.

11. The composition of claim 10, comprising substantially the same amino acid sequence selected from the group consisting of SEQ ID NOS: 4, 8, 10, 12, 14, 20 and 22.

12. A composition of matter, comprising a substantially purified chicken Eph-related protein tyrosine kinase, or functional fragment thereof having substantially the same amino acid sequence of SEQ ID NO: 2.

13. A composition of matter, comprising a substantially purified chicken Eph-related protein tyrosine kinase, or functional fragment thereof having substantially the same amino acid sequence of SEQ ID NO: 6.

14. A method of diagnosing cancer, comprising removing a tissue or cell sample from a subject suspected of having cancer and determining the level of Eph-related protein tyrosine kinase in said sample, wherein a change in the level or activity of a Eph-related protein tyrosine kinase compared to a normal sample indicates the presence of a cancer or correlates with a specific prognosis.

15. The method of claim 13, wherein an increase in said change in the level or activity of a Eph-related protein tyrosine kinase indicates the presence of a cancer or correlates with a specific prognosis.

16. The method of claim 13, wherein a decrease in said change in the level or activity of a Eph-related protein tyrosine kinase indicates the presence of a cancer or correlates with a specific prognosis.

- 5 17. The method of claim 12, wherein said cancer is selected from the group consisting of liver carcinoma, lung carcinoma, breast carcinoma, colon carcinoma and leukemia.

Ceq5 1 MEGPERUMGPELWCCIPALPLILAAVEETIMDTTATAELGNVHP-PSGWEEVSGYDENMTTIRTYQVCNVE-SSQNNWLTAKYIRRP-GAHRH
 Ceq10 1 GVSSRRARPPGSSRSRRGV.S.A.TT.-ET.....A.OP.....Q.R.-AQ.QQ.....F.N.Q-DVQ.VY
 Ceq6 1TR.....TAN.....L.....PN.....L.TF.N.....Y
 Ceq4 1 DRRRIPIAL.CAALGSAGR.SARPGN.VN.L.K.IQG.....ISY.-SH.....I.V.HYTP.....ES.-MD-H.....NW.P.N-S.QK.Y
 Eck 1 MEIQRARACEAL..G.-A.AAAAAQCK.VV.L.FAA.GG.....LT.YGK.DLMQNDMD.-P.YM.S.....MS-GD.D.....NWVY.G-E.E.NN
 Eph 1 MERRW.LGIGIV.LL.-AP.P.GAR.K.V....TSK.QG.....LLD.PKD.S.QOQILNGT-PLYM.D.PMOGRDRTDH....SNW.Y.GEE.S.V.

Ceq5 96 VEMKFSVRDCSSIPNVPG--SCKETFNLYYESDFDSATKTFPNWMPMKVDITIADESFSQVDLGRWMKINTEVRSFGPVSKNGFYLAFOYGGCM
 Ceq10 79 ..L..T....K...KI.....F.....T....SANS.F.....YI.....P.....KLES.-.....K.....L.....L.A..
 Ceq6 68 T..R.T.....L.....T.SVL....KSAF.T.A.YL.....F..L..GF-----
 Ceq8 1 GESQ---FA.I.....T...I.D.I..L.....DV..L..K.....V.A.I
 Ceq4 96 ..L..TL...N...L.L.--T.....M...D.HLA.-.REHQ----FT.I.....T.M...D.I.L.L.....EV.....K.....V.A.V
 Eck 96 F.IN.T....N.F.CGAS--.....A...L.YG.N-QKRL---FT.I....P..ITVSS.FEA.HV.L.V.E..V..LTRK.....I.A.V
 Eph 97 ..LQ.T....K.F.GGA.PLG.....L.M...Q.VGIQ-IRREL---FQ..T.V...Q..TIR..ASGSV.L.V.RC.L.RLRR.L....HNP.A.V

Ceq5 194 SLIAVRVYRKCPRTVIONGAVFTOETLSGAESTSLVAARGTCISNA---EEVDVPIKLYCNGDGEWLVPICRCMCRPGYSEVNGTVCRGCPSTGTFKASQ
 Ceq10 173 ..L..A..K..SNT.AGF.I.P...T...P...I.P...PQ.---V..S..L.....M.V.A.T.AA..PAMKD.Q.OA.GP.....SK..
 Ceq6 141 -----FK...S.V..F.I.P.MT.....T.....P.....M.....T.KA...PEN.-VA..A..A.....
 Ceq8 56 A.VS.....K...LTVR.L.Q.PD.IT..DTS...EV..S.VN.S---.K...-M.GA.....N.L.NA...ERNG--E.OA.KI.YY..LST
 Ceq4 190 A.VS...YFK...FTVK.L.M.PD.VPM-D.Q...EV..S.VNHS---K.EEP.-M..STE.....K.L.NA...ERGF--A.OA.RP.FY...A.
 Eck 190 A.LS...Y.K...ELL.GL.H.P..IA.SDAP..ATVA...VDH.VVPPGGEE.-RMH.AV.....Q.L.OA...K..D--A.OA.SP.F..FEAS
 Eph 193 A.VS.....QR..ETLNGL.Q.PD..P.--PAG..EVA...LPH.RASPRPSGAPRMH.SP...V...H.E....EGGS.EA.VA....SYRMDMD

FIG.1A

2 / 9

```

Cek5 291 DEGCVHCPIINSRTTSEGATNCVRCGYRADAPVDMPCITIPSAPOAVISSVNETSIMLEWTPPRDSSGREDLVNIICKSCGS---CRGACTRCGDNV
Cek10 270 EGP.SP...P...AGA..V.I..S.FF.....A.SA..SV.....RS...N.....FV...SE.Q.A..D..L.V...K.SV---E.RL.S..D...
Cek6 230 AGL.AR..P...SSA.ASPL.A.....F...L..PTAA..SV..G.RN...I.....II...N...ET...D.VT...V..K.RA---D.R..S..D...
Cek9 1 V.....
Cek8 150 VA.AK..PH.YSIW..S.S.T.DR.FF..EN.AAS.....RP.....NL..N.....VN...SA.ONK...D.IS.VV..R..A---EPHS.RS..SG.
Cek4 283 NVK.AK..PH.S.YEDASIN.R.EKN.F.SEK..PS.A..RP.....RN...NI.....VI.D.SW.L.T...K.VTF.....K..G---SSKI.EP.S...
Eck 287 ESP.LE..EHTLPSP.....S.E.EE.FF..PQ..AS.....RP.....HYLTAVGMGAKVE.R...Q.....I...SVT.EQ.WP---ES.E.GP.EAS.
Eph 291 TPH.IT..QQ.TAE.....I.T.ES.H..PGECPQA..GP.....RNLSF.ASG.Q.S.R.E..A.T...Q.VR.SVR.SQ.QGTAQDG.P.QP..VG.

Cek5 388 QFAPRQLGLTEPRIYISDLAHTQYTFEIOAVNGVTDQSPFSPQFASVNITNOAPSAVSIMHQVSRVTSITLSWSQDPQNGVILDELYQYKRL-
Cek10 367 E.V.....R.....KVM..P.....ISSK..YP.H.....VL...PT..LH.S.GN.M...TP.ER...I.....IK.S..QGQ
Cek6 327 E.V.....T.VF..S.W..P.....SNK...P..HV.....T.P.....A.MR.....P.E...I.....R...LS+
Cek9 8 ..E...V...S.VQV.N..RV.....L..EL.SEA..Y.TI.VS.S.SV..IPM.....ATS.....P.....Q.R.FD.AE-
Cek8 248 H.S.Q.N..KTKVS.T.....N.....VW.....SKHN.SOD.AV..TV.....PIALIOAKEI.RH.VA.A.LE..R.....E.VK...DQ-
Cek4 380 R.L...T...NTVTTV.....N.....D.....S.L.TL.R..A.S.....PITVIRKORTSRN.VS..QE.EH...I.....VK...QE-
Eck 384 RYSEPH...RTSVTV...EP.MN...TVE.R...SGLVT--RS.RTASVSI..TE.PK.RL---EGRSTT.LSV...I.PPOQSRVWKEI.VT.RKNGD-
Eph 391 H.S.GARA..T.AVHNG.EPYAN...NVE.Q...SGLGSSGHAST..S.SMGH.ESLSGLSLRL.KKEPRQLE.T.AGSRPRSP.AN.T...HVLNQD-

Cek5 487 SELNSTAVKSPINTVTVQNLKAGTIYVFQVRARTVAGYGRYSGRMXFQMTAEAYQTSVQEKPLLIIGSSAAGINVELIAVVIIVCN-----
Cek10 467 GDGIANT.T.QK.S.RLDG...NAR.MV.....LPTE...TA.DGSTSKTFOE...V..AT...L.V.V..I..A...F----- (RKGMVT
Cek6 445 N.Y...SVAR.Q...ARLEG.RP.MV..V.....K.....C...L.DDD.KSELR.Q...A..A...V..IVSL.A.S...S-----
Cek9 107 D.D..FTLT.E..MA..IL..SP.K.....AV...P.....LMGG.HSEMR.DR...V..ALG..A..VIAAIA..AIL-----
Cek8 347 N.RTYRI..TASRNTDIKG.NPL.S...H.....A...DF..PFE.T.N.VPSP-IIGDGTN.TVLIV.V..S.V.VIILIAF..IS-----
Cek7 1 .....I.....A.....GF.RRFE.E.SPVLAAR-S.D.SQI..I.VV.VTV.VII.AV.IGELISGCCDHCWGASSL-----
Cek4 479 Q.TSY.IIRAKSTN..ISG..PD.T...I.....A.R..TS.R.FE.E.SPDSFESIS.ENSQVAM..AT.A.VAILL.TV..VVL.G-----
Eck 479 -SNSYNVRTEGFS..LDD.APD.T.IV..Q.L.QE.Q.AG.KVHE...LS-----PEGSGNIAY..GV.V.V.L.IVAGVGEFIH-----
Eph 491 ----ERYQWLEPR.LLTE.QPD.T.IVR..ML.PL.P.PF.PDHE.R.SPPVSR-GLTGGEIVAV.FGILL.AAL.IGIL.FRSRRA-----

```

FIG.1B

.PA
 Ceq5+ STYRGPPGLGVRLFV
 Ceq5 575 -----RRRGERADSEYTDKLOHYTSGH-----MTPGMKIYIDPFTYEDPNEAVREFAKEIDISCVKIEQVIGAGEFGEVCSGHLK
 Ceq10 555 EQLLSSPLG).KORNST.P...E...Q.VT-----V.....E.....R.R..
 Ceq6 533 -----KRAYSKEV.S.....ST.R-----GS.....V.F...E.....YK.R..
 Ceq9 195 -----FKSK.RETP...R..Q.I.TR-----GL.V.Y...S.....I.....V.FI...E...S.....F.R..
 Ceq8 434 -----R..SKYSK.KQ.ADEEKHLN-----Q.VRT.V.....Q.....A.I...K...V.....R..
 Ceq7 75 RAVAYPSLIW.C.YSK.KQDPEEKM.FHN..-----IKL..VRT...H.....Q..H.....EA..IT..R.....R..
 Ceq4 565 -----FC.YKSKMGHGTDE.RL.FGN..-----LKL..LRT.V..H.....Q..H.....L.A.NIS.DK.V.....R..
 Eck 560 -----RKNQARQSPEDVYFSK.EQ-----LKPL.T.V..H.....Q..LK.TT..HP...TRQK.....YK.M..
 Eph 574 -----Q.QROQ.HVTAPPMWERTSCAE-----ALCG.SRHTIHLREPWTI.GGWSNEPSR.L.PAWLMVDT...E.....YR.T.R

Ceq5 651 L-PGKREIFVAIKTLKSGYTEKORRDFLSEASIMGQFDHPNVIHLEGVWTKSSPVMIITEFMENGSLDSFLRQNDGQFTVIQLVGMRLRGIAAGMKYLADM
 Ceq10 626 -.R.....V...R.....I.....R.....CA.....L.....SE.
 Ceq6 608 -.Y.....A..S.....I.R.....R.....A.....E.
 Ceq9 268 H-.....YT.....DE...E.....E.....R.....V.....KE...S.L.....R..S..
 Ceq8 505 V-.....C.....A..D.....I.....CK.....Y.....A..K..R.....GS..
 Ceq7 161 -.Q....FP.....V.....G.....I.....K...V..Y...T.KK.....S..
 Ceq4 642 -.S.K..S.....A.....G.....I.....K...V..Y...KH.A.....S..
 Eck 634 TSS..K.VP.....A.....V...G.G.....S.H.I.R...IS.YK.M...Y...A..K..EK..E.S.L.....N..
 Eph 653 -.SQDCKT.....DTSPGG.WWN..R..T.....S..HIL.....RK.I.....AA..A...ERED.LVPG...A..Q..S..N..SNH

FIG.1C

4 / 9

Cek5 750 NYVHRDLAARNILVNSLVCKVSDFGLSREFLEDDTSPTTYSALGGKIPIRWTAPEAIQYRKFTSASDVWSYGIWWEVMSYGERPYWDMTNQDVINAIE
 Cek10 725PA.....S.....A.....S.....S.....S.....S.....V.
 Cek6 707Y.Q.....S.....V.....A.....F.....S.....
 Cek9 367A.N.....G.....C.....V.....S.....S.....D
 Cek8 604 S.....M.V.....-PEAA..TR.....A.....S.....K
 Cek7 260 G.....I.....V.....-PEAA..TR.....AF.....E.....K.V.
 Cek4 741 G.....I.....V.....-PEAA..TR.....S.....A.....L.....E.SF.....K.VD
 Eck 734V.....-PEA.....TS.....S.....F.....T.....ELS.HE.MK..N
 Eph 752Q.C.....T.L.D.--F.G..ETQ.....AH.I..T.....L.F.DK..GE.S..E.MKS..



Cek5 850 QDYRLPPMDCPNALHQLMLDCWQKDRNHRPKFGQIVNTLDMIRNPNSLKAMAPLSSGVNLPLLDRTIPDYTSNTVDEWLDAIKMSQYKESFASAGFT
 Cek10 825T.....VR...L...A.....L...AA...VI.SVQ...SQ.....V....T.T..GD.....GR...N.VN...A
 Cek6 807A.....T.RLAE.....A.....TV.TITAVPSQ.....S...F.A.TS.ED..S.V....RDN.L.....
 Cek9 467P...TV..L.....VQ...E...SA.....K.SA...TGTG..RPSQ...SNRP..FP.LSNAH.....GR...N.DQ..LI
 Cek8 702 EG.....I.....E.SD.....M..L.....RTGSE..RPSTA...PSS.EFSAVVS.SD..Q....ER..DN.TA..Y.
 Cek7 358 EG...S...A.Y.....S...DE..SM..L...S...TIVNA..R.SNL.VEHSVPVSGAVRS.G...E....GR.T.I.MEN.YS
 Cek4 839 EG.....A.Y.....N...E...SI...L...S...IITNAAARPSNL...QSN.I.SA.R.AGD..NGFRTG.C.GI.TGVEYS
 Eck 832 DGF...T....S.IY...MQ...QE.AR...AD..SI...L..A.D...TL.DFDP.R.SIR.PSTSGSEGVP.R..S...ES...Q..T.H.MA..Y.
 Eph 849 DG.....V...AP.YE..KN..AY..AR..H.QKLOAH.EQLLA..H..RTI.NFDPR.T.R.PSLSGS.GIPYR..S...ES.R.KR.IILH.H...LD

Cek5 950 TFDIVSQMTVEDILRVGVTLAGHQKKILNSIQVMRAQNIQISVEV
 Cek10 924 S..L.A...A..L..I.....S...D..L....TLP.Q.
 Cek6 907 SLQL.A...S..L..I.....S.V..S.SPTSMA
 Cek9 567 ...VI.R..L..IQ.I.I..V.....L.KVHL..LEP...
 Cek8 802 .LEA.VH.NQD.LA.I.I.AIT..N...S.V.A..S..Q.MHGRM.PV
 Cek7 458 SM.S.A.V.L..-----ESPCE.WSLTLHPLFPTGY.T
 Cek4 939 SC.TIARISTD.MKK...VV.P...VS..KILETHTKNSPVV
 Eck 932 AIEK.V...ND...K.I..R.P....R.AY.LLGLKD.V.TVGIP
 Eph 949 .MEC.LEL.A..LTOM.I..P.....R..C..GFKD

FIG.ID

5/9

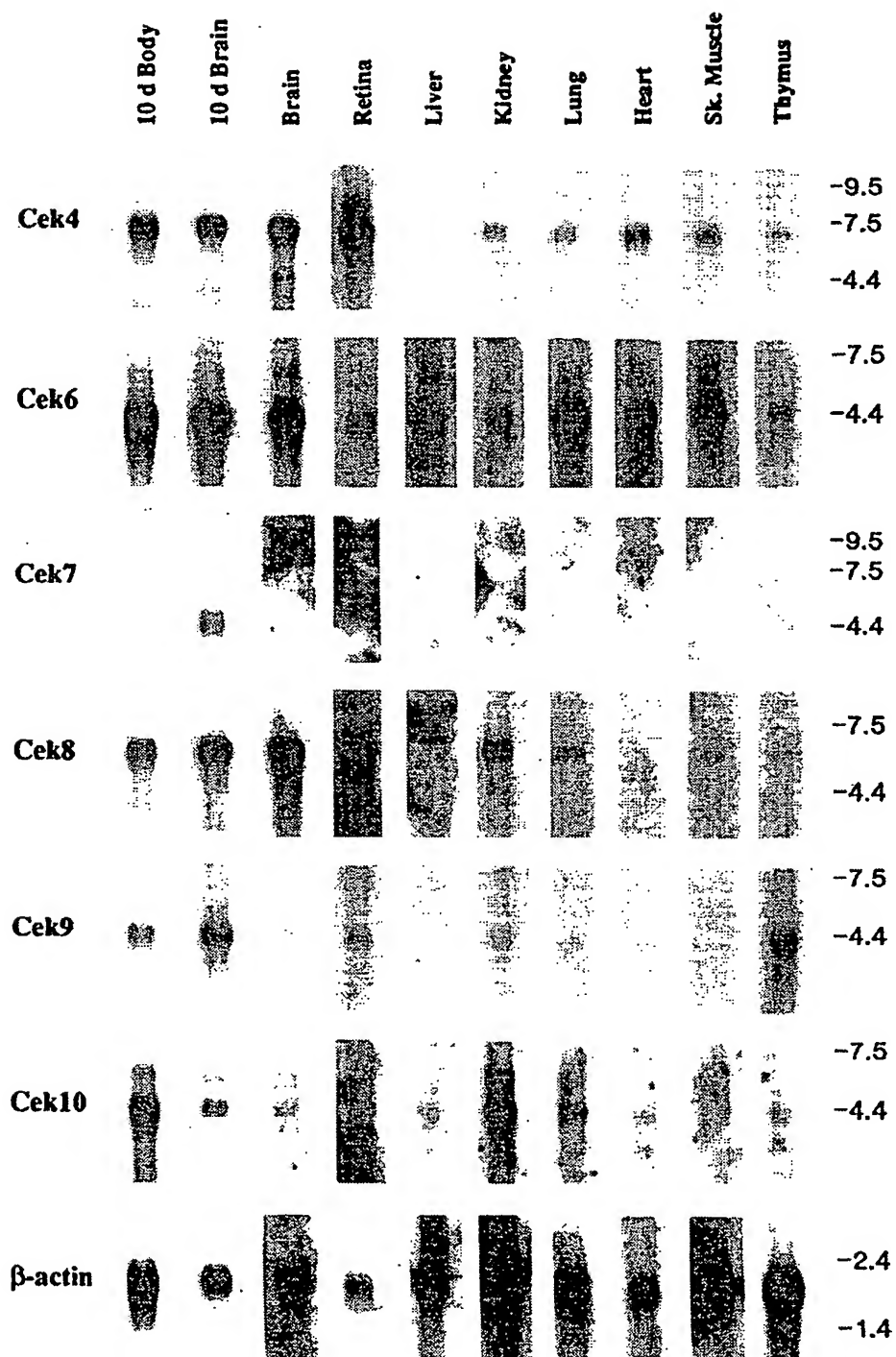


FIG. 2
SUBSTITUTE SHEET (RULE 26)

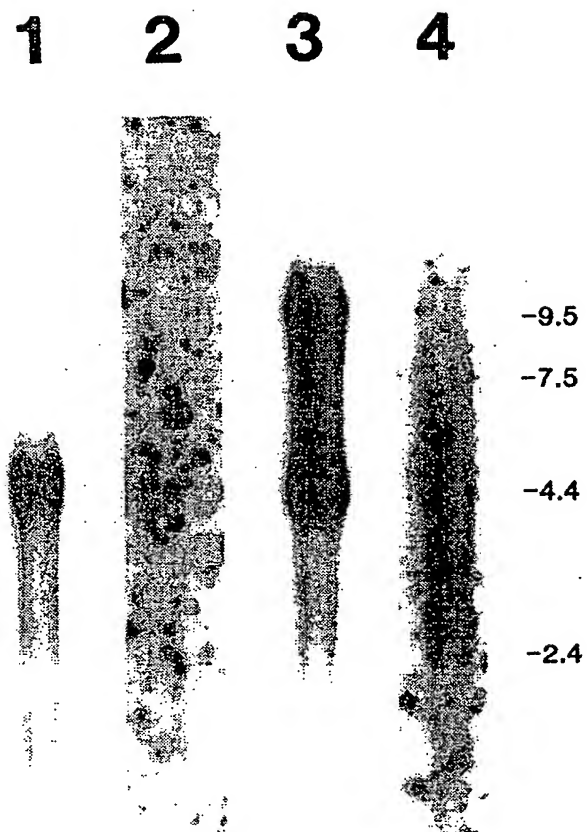


FIG. 3

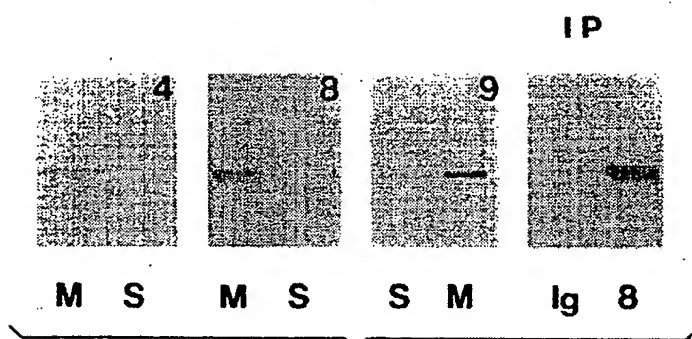


FIG.4

8/9

FIG. 5A

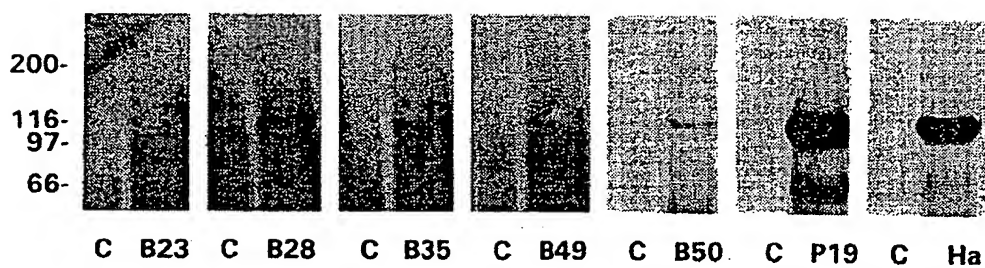


FIG. 5B

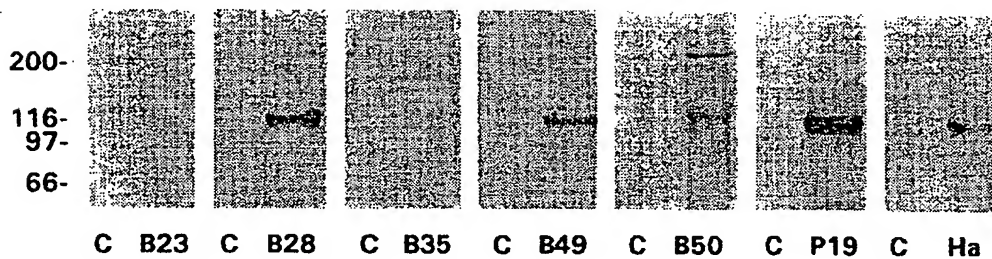


FIG. 5C

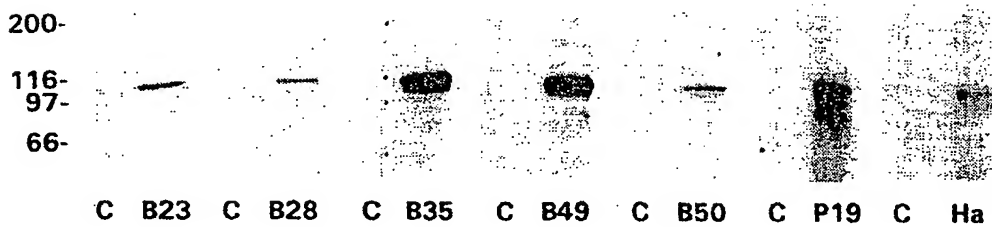
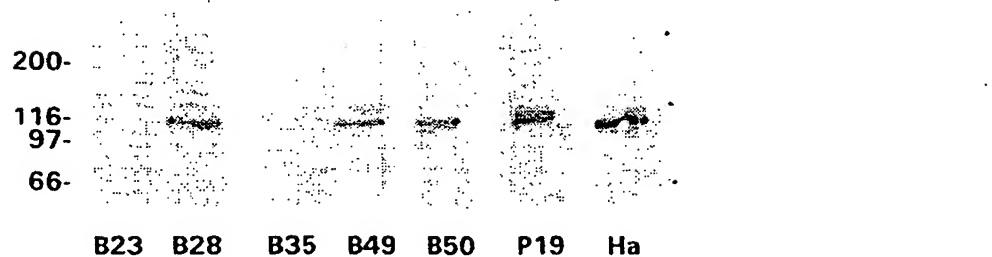


FIG. 5D



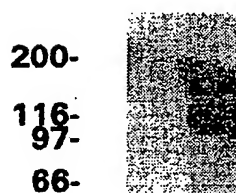
SUBSTITUTE SHEET (RULE 26)

FIG. 6A



10d Ad LMH

FIG. 6C



CEF RSV

FIG. 6D



CEF RSV



FIG. 6B

FIG. 6E

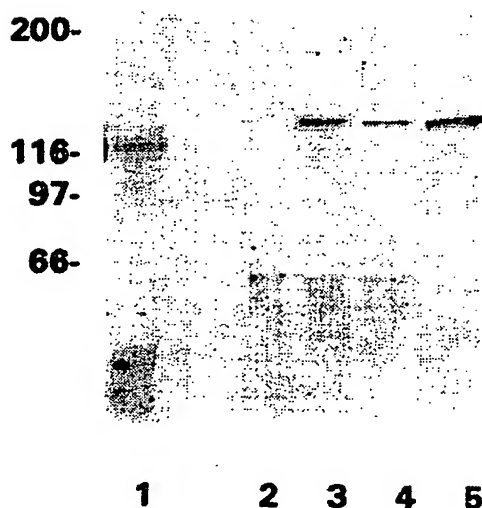
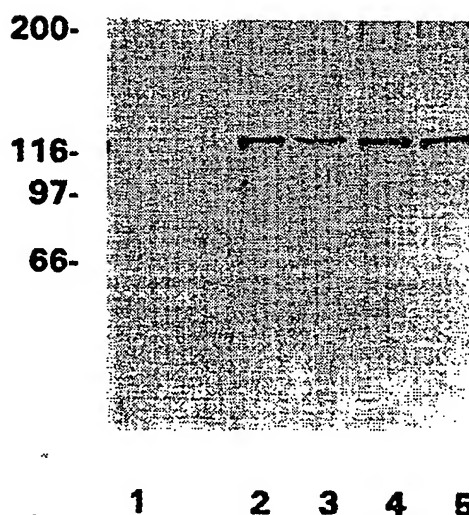


FIG. 6F



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US94/10140

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12N 15/00, 9/00

US CL :435/240.2, 252.3, 320.1, 194; 536/23.2, 23.5

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/240.2, 252.3, 320.1, 194; 536/23.2, 23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
noneElectronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Dialog, APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Oncogene, Volume 320, issued 1992, Gilardi-Hebenstreit et al., "An Eph-related receptor protein tyrosine kinase gene segmentally expressed in the developing mouse hindbrain", pages 2499-2507, see entire document.	1-9
Y	Cell Regulation, Volume 2, issued July 1991, E. B. Pasquale, "Identification of chicken embryo kinase 5, a developmentally regulated receptor-type tyrosine kinase of the Eph family", pages 523-534, see entire document.	1-9

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	* T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principles or theory underlying the invention
* A		document defining the general state of the art which is not considered to be of particular relevance
* E		earlier document published on or after the international filing date
* L		document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
* O		document referring to an oral disclosure, use, exhibition or other means
* P		document published prior to the international filing date but later than the priority date claimed
	* X	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
	* Y	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
	* &	document member of the same patent family

Date of the actual completion of the international search

22 NOVEMBER 1994

Date of mailing of the international search report

10 JAN 1995

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

KEITH D. HENDRICKS

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/10140

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Proceedings of the National Academy of Sciences USA, Volume 89, issued March 1992, Wicks et al., "Molecular cloning of HEK, the gene encoding a receptor tyrosine kinase expressed by human lymphoid tumor cell lines", pages 1611-1615, see entire document.	1-9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US94/10140

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

- I. Claims 1-9 drawn to the DNA, vector, host cell and method of making protein.
- II. Claims 10-13, drawn to the enzyme.
- III. Claims 14-17, drawn to a method of using the enzyme of Group II.

The inventions listed as Groups I-III do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The DNA Group I and the protein of Group II are not chemically related, and they are independent and distinct inventive concepts. The DNA possesses utility other than encoding the protein, such as in a hybridization assay or as a probe. The DNA of Group I is not related to the method of Group III for similar reasons, as the method of Group III does not involve the DNA.

Groups II and III are related as a second product and method of use. The method may be used with a materially different enzyme, such as one from another source.

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.